



# Patterns of Science-Technology Linkage



*Research and  
Innovation*



EUROPEAN COMMISSION  
Directorate-General for Research and Innovation

E-mail: [rtd-innovation-papers-studies@ec.europa.eu](mailto:rtd-innovation-papers-studies@ec.europa.eu)  
[RTD-PUBLICATIONS@ec.europa.eu](mailto:RTD-PUBLICATIONS@ec.europa.eu)  
[Carmen.Marcus@ec.europa.eu](mailto:Carmen.Marcus@ec.europa.eu)

Contact: Carmen Marcus and Pierre Vigier (Head of unit)

European Commission  
B-1049 Brussels

# Patterns of Science-Technology Linkage

Authors of the study

Julie Callaert, Jan-Bart Vervenne, Bart Van Looy, Tom Magermans, Xiaoyan Song and Wouter Jeuris

Research Division INCENTIM

*MSI, Faculty of Business & Economics, K.U. Leuven  
Università Commerciale Luigi Bocconi, KITES*

This report is part of the study *Measurement and analysis of knowledge and R&D exploitation flows, assessed by patent and licensing data*, carried out by K.U Leuven (coordinator), Bocconi University and Technopolis Vienna under the coordination and guidance of the European Commission, Directorate-General for Research and Innovation, Directorate Research and Innovation, Economic analysis and indicators Unit.

EUROPE DIRECT is a service to help you find answers  
to your questions about the European Union

Freephone number (\*):

**00 800 6 7 8 9 10 11**

(\*) Certain mobile telephone operators do not allow access to 00 800 numbers or these calls may be billed

## LEGAL NOTICE

Neither the European Commission nor any person acting on behalf of the Commission is responsible for the use which might be made of the following information.

The views expressed in this publication, as well as the information included in it, do not necessarily reflect the opinion or position of the European Commission and in no way commit the institution.

More information on the European Union is available on the Internet (<http://europa.eu>).

Cataloguing data can be found at the end of this publication.

Luxembourg: Publications Office of the European Union, 2014

ISBN 978-92-79-34793-1

doi:10.2777/55249

© European Union, 2014

Reproduction is authorised provided the source is acknowledged.

Cover images: earth, © #2520287, 2011. *Source:* Shutterstock.com; bottom globe, © PaulPaladin #11389806, 2012. *Source:* Fotolia.com

## Table of Contents

1. BACKGROUND AND RESEARCH OBJECTIVE .....	4
2. DATA AND METHODOLOGY .....	7
2.1. Data sources .....	7
2.2. Data enhancements .....	8
3. ANALYSES AND RESULTS .....	14
3.1. Supply versus absorption of science in national-technology development.....	14
3.2. Mapping of cross-country citation flows between science and technology (based on patents citing scientific literature).....	17
3.3. Analysis of the relation between national technological performance and science-technology linkage patterns on a national level.....	26
4. SUMMARY CONCLUSIONS & DIRECTIONS FOR FOLLOW-UP STUDIES.....	33
REFERENCES .....	35
APPENDICES & ANNEXES .....	38
Appendix 1: Patstat relational diagram .....	39
Appendix 2: Fraunhofer Technology classification (35 fields) (Schmoch, 2008) .....	40
Appendix 3: Cross-country science-technology linkage matrix, based on absolute numbers of citations.....	44
ANNEX 1a: Cross-country citation matrices for corporate patents citing scientific documents by FhG-35 technology class .....	45
ANNEX 1b: Comparison of within-country citation intensity to the average citation intensity with foreign countries, by FhG-35 technology class .....	80
ANNEX 2: Detailed account of WoS-NPR matching approach.....	89
ANNEX 3: Supply versus absorption of science per citing country, by patent office .....	98

# 1. BACKGROUND AND RESEARCH OBJECTIVE

This report on Patterns of Science-Technology Linkage is part of the project's fourth work package and aims to analyze evidence on knowledge flows or interactions between scientific and technological activities and actors, with a special focus on cross-country patterns.

The relevance of mapping and monitoring science-technology interactions is implied by today's concepts of innovation systems, in which the dynamic interplay between a variety of actors (knowledge-generating institutes, firms, government,...) plays a crucial role for understanding the dynamics behind innovative performance, growth and competitiveness of nations (see e.g. Freeman, 1987, 1994; Adams, 1990; Lundvall, 1992; Nelson, 1993; Mowery and Nelson, 1999; Baumol, 2002; Veugelers et al., 2012). Indicators signaling interactions between scientific and technological activities are highly relevant in this respect. Researchers and policy makers have since several decades recognized patents as valid and reliable indicators of technology development and innovation. The availability of information in patent databases (including applicants and inventor names and addresses, affiliations, sectors of activity, references to prior art patent and non-patent literature,...) allows for thorough mappings of involved actors and interactions on several levels of analysis. More specifically, for measuring science-technology linkage, indicators derived from non-patent references (NPRs) in patent documents have become popular. In spite of some discussion about their actual meaning (Alcacer & Gittelman, 2006; Breschi & Lissoni, 2001; Nelson, 2009), scientific references in patents are in any case indicative of relatedness or closeness between the developed technology and the cited science (Callaert 2006; Meyer, 2000a; Tijssen et al. 2000; Van Looy et al. 2007). The presence of scientific references in the front page section of a patented invention indeed signals the relevancy of these references for assessing and qualifying the claims of the invention, so that derived indicators provide useful information on science-technology relatedness or vicinity, at least if their presence displays sufficient levels of occurrence (Callaert et al. 2006). Given the widespread and consistent availability of reliable and comprehensive patent databases, these indicators bear the potential to provide a broad and systematic view on science-technology interactions. In spite of some interpretational pitfalls, they entail a large potential for addressing policy relevant questions on national (and supra-national) innovation systems. The analyses brought forward in this report fulfill at least part of the implied potential.

One of the cornerstone motivations for the current project is the measurement and monitoring of knowledge and R&D flows by using patent-related data. Building on the citation data extracted in previous work packages, it is now feasible to map and

study 'flows' or linkages between science and technology. For analyzing flows, we rely primarily on the reference data available in patent documents, i.e. the information that is available on prior art cited in patent documents. Such prior art can pertain to both patent documents and non-patent documents (non-patent references or NPR's). Previous studies have proven that at least half of these NPR's can be considered 'scientific' (Callaert et al., 2006). For the purpose of mapping flows or linkage between science and technology, prior art can be informative. Their relevance stems from the different patent-related indicators that can be used for assessing science-technology linkage (Van Looy et al., 2002). An overview of these indicators is presented in Table 1.

1. Patents	2. Citations in patents
(1.1) Number of patents assigned to or knowledge-generating institutes	(2.1) Non Patent References
(1.2) Number of co-patents assigned to knowledge-generating institutes and co-owned with firms	(2.2) Patent References

**Table 1: Overview of patent-related indicators for assessing science-technology linkage (adapted from Van Looy et al. 2002).**

This report on Patterns of Science-Technology Linkage will focus on indicator (2.1). Given the presence of scientific (non-patent) references in patent documents, science-technology linkage can be assessed by considering the occurrence of scientific articles in patent documents and by identifying the source of origin (affiliations) of the cited scientific work.

The geographic (national-level) projection of science-technology patterns - performed in the project's previous work package 1.3 (deliverables 1.5 and 1.7) allows for studying cross-country patterns of science-technology linkage. This is highly useful for addressing questions on the sources of national technological development, as drivers behind a nation's innovative performance. Several questions are relevant in this respect. The following research questions will guide the analyses.

### **1. Do supply and absorption of scientific resources for technology development coincide on a national level?**

This analytical section considers the demand and supply of science. We will analyze whether countries with a large 'supply' of science also display high levels of absorption of science in their technology development. More specifically, the developed indicators allow assessing whether national scientific performance is translated into higher science intensity of the national patent portfolio.

## **2. What are patterns of cross-country citation flows between science and patents?**

The previous research question addresses national-level relations between supply and absorption of science, but makes abstraction of where science is sourced: from the home country or abroad? A next analytical section will therefore focus on cross-country citation flows. The main question in this section relates to the presence of a home bias or geographic proximity effects in the citations from a country's corporate patents towards science.

## **3. What are the relations between national-level characteristics of science-technology linkage patterns and national-level technological performance?**

An assessment of the relation between science-technology linkage patterns (based on the answers to the previous research questions) and national technological performance will conclude this study.

In what follows, the data and indicators used will be outlined, after which analytical results will be reported on. The report ends with a conclusion and directions for follow-up studies.



## 2. DATA AND METHODOLOGY

As outlined above, the availability of scientific (non-patent) references in patent documents allows for assessing science-technology linkage by considering the occurrence of scientific articles in patent documents and by identifying the source of origin (affiliations) of the cited scientific work. In what follows, we shed more light on the data sources used and the construction of relevant indicators.

### 2.1. Data sources

All patent data are extracted from the EPO PATSTAT Database (Worldwide PATent STATistical Database). It is a patent statistics database, held by the European Patent Office (EPO) and developed in cooperation with the World Intellectual Property Organisation (WIPO), the OECD and Eurostat. In a relational format, PATSTAT contains over 20 tables including detailed bibliographic data, applicant and inventor information, citations, patent and non-patent references,... (for the full relational scheme: see the PATSTAT relational diagram in appendix 1). For the indicators used throughout our analyses, the relevant tables are:

- 'CITATION' table: contains references to prior art, mainly patents.
- 'PERSON' table: detailed information about inventors and applicants (name, address information).
- 'NPL\_PUBLN' table: contains the 'NPL\_BIBLIO' field, which is the cited non-patent reference. This text string will be matched to the scientific citations in the Web of Science database.

The source of publication data (scientific articles) is the ISI / Thompson Reuters Web of Science database (SCIE, SSCI, AHCI) and the ISI Proceeding databases (CPCI-S and CPCI-SSH). These relational databases cover scientific journals and proceedings and contain detailed information on authors, affiliations, disciplines, etc...<sup>1</sup>

For analyzing cross-country flows between citing and cited documents, author and applicant countries were identified for scientific publications and patents respectively. Both the PATSTAT and the Web of Science & Proceedings databases contain this information. For documents allocated to more than 1 country (due to multiple applicants or authors), a full fractional counting scheme is used<sup>2</sup>.

Besides national allocation of patents and scientific publications, some more challenging data enhancements were required for obtaining the indicators on which this report is based. The most crucial enhancements concern the sector allocation of patent applicants and the identification of cited non-patent references (matching

---

<sup>1</sup> Note that the available scientific databases cover publications from 1991 onwards.

<sup>2</sup> E.g. for a patent that is applied for by 1 French applicant and by 2 German applicants, an equal flow weight of 1 is assigned to France and 1 to Germany.

scientific non-patent references to the Web of Science database). These enhancements will be outlined briefly in the following section 2.2. For more detailed accounts of these enhancements, we can refer to previous project reports.

## **2.2. Data enhancements**

### ***2.2.1. Applicant sector allocation***

For the identification of academic and corporate patents, the sector allocation methodology (identification of patent applicants as being universities, companies, government agencies,...) developed at ECOOM KU Leuven (Van Looy et al., 2006) and further refined by Du Plessis et al. (2009) is used. This methodology is a combination of a rule-based and a case-based logic. The rule-based logic starts from information incorporated in the applicant names that provide clues on 'sector' membership, which can then be translated into a set of rules for the automated allocation of sector codes. The case-based logic is applied as a complement for records lacking clues or with the simultaneous presence of several clues that suggest different sectors. Quality levels, both in terms of completeness and accuracy, of at least 99% are obtained: less than 1% of the patent volume fails to obtain a sector code; the patent volume accurately assigned to a particular sector exceeds 99%.

### ***2.2.2. Identification of NPRs: characterization of scientific NPRs***

Non-patent references in patent documents are generally considered and treated as 'scientific references'. However, not all non-patent references are 'scientific' in the strict sense of the word. Callaert et al. (2006) analyzed a sample of 10,000 EPO and USPTO NPR's and found that almost half of them referred to sources other than journal articles. These concerned conference proceedings (which are scientific documents as well), but also: industry-related documents (company catalogs, manuals), newspapers, reference books or databases, gene/plant bank records and patent-related documents. Hence, non-patent references contain some noise that needs to be eliminated if one wants to analyze links between actual scientific documents and patents. Callaert et al. (2011) developed an algorithm – based on supervised machine learning – for automated classification of NPRs' into scientific and non-scientific sources. A detailed description of this procedure can be found in Callaert et al. (2011).

The resulting filter on real scientific NPR's is important for our purpose of matching non-patent references to the Web of Science. Being able to eliminate noise (non-scientific references) in the large set of non-patent references reduces the amount of 'matchable' references considerably, which in turn implies a higher efficiency of

the matching procedures. The identification algorithm from Callaert et al. (2011) was applied to the PATSTAT September 2010 edition. Table reports the reduction in non-patent references if only scientific ones are considered, broken down by application year. It can be seen that the scientific NPR's represent little more than 52% of all non-patent references.

Application year	#NPRs	# Scientific NPRs	% of Scientific NPRs
<b>1993</b>	352.816	184.116	52,18%
<b>1994</b>	460.503	263.780	57,28%
<b>1995</b>	723.368	453.446	62,69%
<b>1996</b>	532.667	283.248	53,18%
<b>1997</b>	674.315	366.002	54,28%
<b>1998</b>	703.119	376.749	53,58%
<b>1999</b>	812.603	438.851	54,01%
<b>2000</b>	965.667	506.236	52,42%
<b>2001</b>	1.027.899	531.232	51,68%
<b>2002</b>	1.107.040	573.208	51,78%
<b>2003</b>	1.066.291	536.976	50,36%
<b>2004</b>	1.105.543	536.949	48,57%
<b>2005</b>	1.026.694	491.300	47,85%
<b>2006</b>	836.612	393.167	47,00%
<b>2007</b>	614.495	290.909	47,34%
<b>2008</b>	378.668	196.367	51,86%
<b>2009</b>	115.982	76.599	66,04%
<b>Total</b>	<b>12.504.282</b>	<b>6.499.135</b>	<b>51,98%</b>

**Table 2a: shares of scientific NPRs in PATSTAT (September 2010 edition) NPR population per application year**

Besides lowering the efficiency of matching algorithms for identifying scientific references, ignoring the distinction between scientific and non-scientific NPRs may result in a distorted picture when developing indicators of science-technology interaction (Callaert et al., 2011). To illustrate this, Table compares the shares of applications with NPR's versus the share of applications with *scientific* NPR's (SNPR's) for OECD countries, Taiwan and China for the period 1993-2000. It can be seen that this very basic indicator of national science-technology relatedness differs greatly, depending on whether or not non-scientific NPR's are filtered out. Not only are national shares themselves reduced, but the ranks of the countries on the measured indicator shift as well (e.g. US goes from 5<sup>th</sup> to 2<sup>nd</sup> place).

<b>Ctry</b>	<b>Applns</b>	<b>Applns with NPRs</b>	<b>% Applns with NPRs</b>	<b>Rank based on NPR %</b>	<b>Applns with SNPRs</b>	<b>% Applns with SNPRs</b>	<b>Rank based on SNPR %</b>
<b>US</b>	3.462.027	1.017.472	29,39%	5	422.529	12,20%	2
<b>JP</b>	1.970.589	452.344	22,95%	13	146.677	7,44%	13
<b>DE</b>	1.736.392	341.816	19,69%	19	94.239	5,43%	23
<b>KR</b>	983.583	44.321	4,51%	34	14.901	1,51%	34
<b>CN</b>	789.259	11.569	1,47%	36	3.347	0,42%	36
<b>FR</b>	555.564	135.588	24,41%	10	44.777	8,06%	9
<b>GB</b>	351.257	89.411	25,45%	8	30.292	8,62%	7
<b>NL</b>	297.962	80.950	27,17%	7	23.573	7,91%	12
<b>CH</b>	295.624	59.312	20,06%	18	18.482	6,25%	16
<b>SE</b>	245.418	51.934	21,16%	16	13.597	5,54%	21
<b>IT</b>	242.103	35.195	14,54%	23	9.516	3,93%	27
<b>TW</b>	229.571	8.305	3,62%	35	3.572	1,56%	33
<b>CA</b>	155.256	37.976	24,46%	9	16.042	10,33%	5
<b>FI</b>	135.008	29.213	21,64%	15	7.347	5,44%	22
<b>AT</b>	81.355	13.651	16,78%	22	3.830	4,71%	26
<b>DK</b>	74.773	17.346	23,20%	12	6.336	8,47%	8
<b>BE</b>	69.614	16.165	23,22%	11	5.595	8,04%	10
<b>ES</b>	69.225	9.794	14,15%	25	3.320	4,80%	25
<b>AU</b>	61.269	20.445	33,37%	2	5.637	9,20%	6
<b>IL</b>	47.804	14.995	31,37%	4	5.505	11,52%	3
<b>NO</b>	37.883	6.864	18,12%	20	1.903	5,02%	24
<b>IE</b>	21.950	4.614	21,02%	17	1.448	6,60%	15
<b>LU</b>	13.255	2.925	22,07%	14	775	5,85%	19
<b>CZ</b>	13.028	725	5,56%	32	208	1,60%	32
<b>NZ</b>	10.199	3.335	32,70%	3	1.085	10,64%	4
<b>PL</b>	9.644	437	4,53%	33	162	1,68%	31
<b>HU</b>	9.211	1.300	14,11%	26	564	6,12%	17
<b>TR</b>	7.871	766	9,73%	30	71	0,90%	35
<b>SI</b>	3.698	652	17,63%	21	245	6,63%	14
<b>PT</b>	3.137	455	14,50%	24	179	5,71%	20
<b>MX</b>	3.117	365	11,71%	28	122	3,91%	28
<b>IS</b>	1.713	582	33,98%	1	232	13,54%	1
<b>SK</b>	1.574	155	9,85%	29	35	2,22%	30
<b>GR</b>	1.326	364	27,45%	6	106	7,99%	11
<b>EE</b>	962	83	8,63%	31	32	3,33%	29
<b>CL</b>	758	98	12,93%	27	45	5,94%	18

**Table 2b: Amounts and shares of patent applications with NPRs and scientific NPRs for OECD countries, Taiwan and China (all patent bureaus, application years 1993-2009).**

### ***2.2.3. Matching NPRs to the Web of Science***

For being able to map cross- country flows, a country-level allocation of the cited NPRs is required. This means that the specific cited article needs to be identified, as national allocation is based on affiliation address of the author(s). Therefore, a major challenge for the indicator development implied the matching of non-patent references from the PATSTAT database to scientific records from the Web of Science database. The large scale of the data covered (matching millions of NPRs in Patstat to millions of records in the Web of Science) implied that an automated process was required. In the context of the current project, several methods were experimented with, all based on text mining applications. Detailed reports of these methods are available, and reveal that the precision of matching approach levels of 99%. The methodology which was ultimately used (see Methodological Note March 2012) is one where the matching is done field-by-field: every parsed field from an article in the Web of Science database is matched separately to the set of NPR strings. If enough fields from a WoS article are retrieved in an NPR text string, then there is a match between the NPR and WoS article.

In what follows, we present a brief summary of the different steps from the methodology. Annex 2 in this report provides a closer look at figures about underlying data and matching results. More details on the methodology in general can be found in the March 2012 Methodological note (Deliv. 1.5)<sup>3</sup>.

The matching approach consists of the following steps:

Step 1: Calculation of field-based match scores for all NPR-WoS pairs within a priori defined year-based filter (+/- 2 years)

The method starts from match scores, calculated for a selection of seven fields that – in varying combinations – allow for the unique identification of individual scientific source documents:

- PY (publication year),
- AUTHOR\_LN (last name of first author),
- SO (journal title),
- VL (volume),
- ISSUE (issue),
- BP (beginning page),
- TI (article title)

---

<sup>3</sup> Note that the March 2012 Methodological report and its key figures covered patents with application years 2001, 2003, 2005 and 2007; whereas the analyses and key figures pertaining to this report and its annex consider all application years from 1993-2009.

Each of these fields is available in the Web of Science, in parsed format.

Match scores between a WoS-field and an NPR string are calculated as: *the number of distinct terms in the WOS field that occur also in the NPR text-string, divided by the number of distinct terms in the WOS-field.*

#### Step 2: Download of NPR-WoS pairs with match scores above specified thresholds

Because of the large size of both the NPR dataset and WoS dataset, it is practically impossible to store all matching scores for all NPR-WoS combinations. Therefore, before downloading resulting matching scores of NPR-WoS combinations for validation and classification purposes, two filters were set to reduce the amount of data to be processed. The filters are defined to eliminate all combinations that are definitely not a match and to only store matching scores of potentially correct matches. The definition of the filters is based on the fact that it requires multiple field matches to results in an overall NPR-WoS match, and that the value of some fields is more important in the matching process compared to other fields (for more details, we refer to Annex 2 and to the March 2012 Methodological Note).

#### Step 3: Exploration and validation of additional filters on match scores to detect 'correct' matches (precision)

The thresholds from step 2 were set sufficiently low to avoid a priori elimination of correctly matched pairs, i.e. to assure recall for those NPRs that are matchable. This implies that the resulting matches with the remaining NPRs will contain a considerable proportion of false positives. In a third step therefore, the match scores for the matched WoS-NPR pairs were evaluated and validated for being able to identify thresholds beyond which the resulting pairs are certain matches, with the aim of eliminating false positives and assuring precision. This validation was performed on the NPR sets from application years 2003 and 2007. Several filters were tested, using different combinations of match scores for the considered fields. For each filter, random sets of resulting pairs ( $N = 100$  à  $300$ ) were manually validated and – based on the results of these quality controls – a decision was made on whether or not to withhold the filter. The latter decision was based primarily on the precision of the matching results, while at the same time, the volume of retrieved matches needed to be sufficiently high for withholding the filter. This step resulted in the definition of 4 filters delineating certain thresholds for different combinations of match scores for selected fields (for more details about the selected filters, we refer to Annex 2 and to the March 2012 Methodological Note).

Step 4: Extraction of the 'correct' NPR-WoS pairs, falling within the criteria specified in step 3.

Based on the matching criteria, resulting matches were extracted and validation and quality controls were performed. Table 3 presents a summary of the resulting matched NPR volumes and the ratio of coverage by application year.

APPLICATION YEAR	TOTAL # NPRs	MATCHABLE NPRs AFTER STEP 2	MATCHED NPRs AFTER STEP 3	% MATCHED NPRs (/MATCHABLE ONES)
1993	352.816	46.426	31.738	68%
1994	460.503	85.839	63.445	74%
1995	723.368	157.986	121.589	77%
1996	532.667	142.728	105.283	74%
1997	674.315	199.290	148.210	74%
1998	703.119	232.285	172.602	74%
1999	812.603	292.806	216.936	74%
2000	965.667	371.207	266.506	72%
2001	1.027.899	414.389	291.999	70%
2002	1.107.040	479.656	340.019	71%
2003	1.066.291	462.135	314.649	68%
2004	1.105.543	486.554	319.131	66%
2005	1.026.694	452.833	298.568	66%
2006	836.612	367.681	241.328	66%
2007	614.495	271.395	181.002	67%
2008	378.668	181.931	125.196	69%
2009	115.982	76.809	53.591	70%

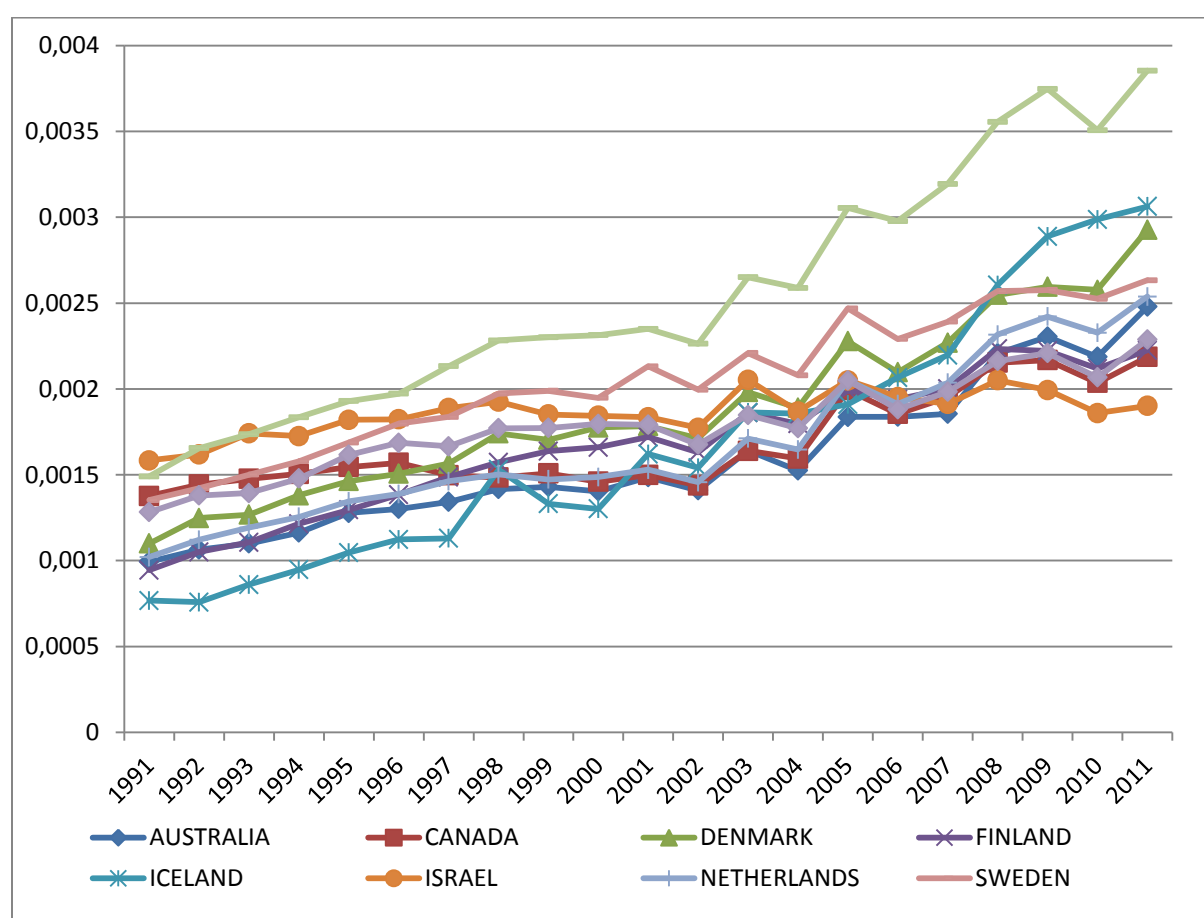
**Table 3: Summary table: NPR volumes and coverage**

The final column in table 3 shows the proportions of NPRs for which the matching was successful, i.e. the proportions of NPRs for which a unique WoS source document was identified (the denominator being 'matchable references', i.e. withheld references that fulfill the defined matching criteria from step 2). It can be seen that between 66% and 70% of matchable NPRs are covered with the defined filters, implying that the developed matching methodology outperforms previous efforts in this respect (cf. Methodological report March 2012). These matched NPR-WoS pairs served as the basis for the indicators and analyses developed in this report, which will be outlined throughout the following sections.

## 3. ANALYSES AND RESULTS

### 3.1. Supply versus absorption of science in national-technology development

This section explores whether the supply of science and the absorption of science in technology development coincide on a national level. In order to quantify this, countries are mapped on two axes: one representing supply (horizontal axis) and one representing 'absorption' (vertical axis). Supply is measured by national-level scientific output per capita<sup>4</sup>. With regard to this measure, the top 10 countries in terms of supply (<OECD, EU-27 and BRIC) are represented in Figure 1 (evolution in scientific publications per capita in the Web of Science, 1991-2011).



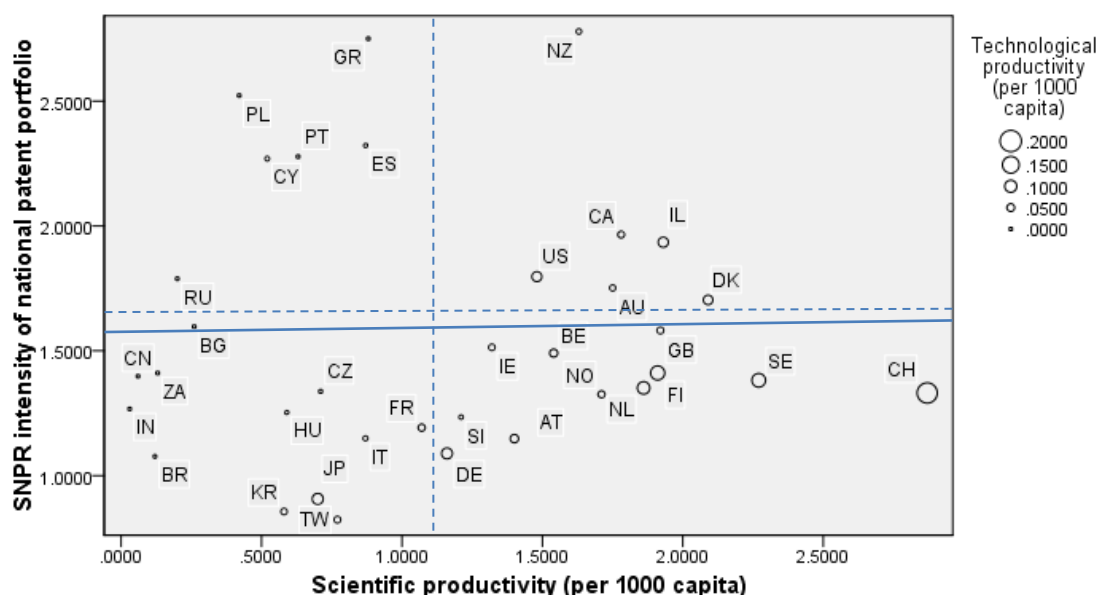
**Figure 1: yearly rates of scientific publications per capita in the Web of Science for the top 10 countries in OECD, EU-27 and BRICS (1991-2011).**

<sup>4</sup> Population levels for OECD countries and China were drawn from the OECD Stats website, the World Bank Indicators website and Census.gov.



The absorption of science in national technology development is depicted by the science intensity of the national patent portfolio, i.e. the average number of scientific references per patent.

Figure 2 maps countries in terms of science supply (X axis) versus absorption of science (Y-axis)<sup>5</sup>. The size of the country dots represents the country's technological productivity. The fitted line shows a slight positive relation between national supply of science and absorption of science in corporate technology development. At the same time, it is clear that there are discrepancies between the science 'supply' and the corporate demand side of science. In order to evaluate these discrepancies, it is useful to distinguish between countries above the fitted line and those below the fitted line. Those above the fitted line show a higher rate of absorption, relative to their supply. They can be suspected to import science from abroad. Countries below the fitted line are stronger suppliers of science, compared to their extent of absorption of science in their national technology base.



**Figure 2: Supply (1991-2011) versus absorption of science per citing country (2000-2009).**

Looking at the graph in more detail, countries in the upper right section are highly productive in the supply of science and at the same time have a high science absorption rate in their corporate technology. It concerns countries like the United States, Israel, Canada and New Zealand. Technological productivity of these

<sup>5</sup> Figure 2 aggregates EPO, USPTO and PCT patents. Annex 3 provides the graphs for each office separately. The pattern for USPTO patents is most similar to what is presented in the aggregate figure. Both EPO and PCT patents have markedly lower science intensities, causing shifts of countries between the upper and lower quadrants (see e.g. the position of US which moves from the upper to the lower right quadrant).

countries is shown to be in the middle higher range. These figures may be indicative of these countries representing effective innovation systems that seek to exploit knowledge through close science-technology interactions. It remains to be seen whether these countries 'absorb' their own science, or whether they rely on foreign (cross-border) science. More light on this question is shed in a following analytical section.

Countries in the lower right section show high levels of scientific output, but their patents are comparatively less science intensive. Scandinavian countries and Switzerland are situated here, as well as the Netherlands, Ireland and Germany. It can be noted that this quadrant gathers the technologically most productive countries, even more so than the upper right quadrant. This suggests that technological productivity is driven primarily by national scientific productivity, and that the relation between technological productivity and science intensity of patents is subordinate to this.

The upper left section of Figure 2 represents countries with a relatively small science base, that are at the same time characterized by a large absorption of science in their corporate technology development. Except for Russia (which is much closer to the lower left quadrant), this quadrant is occupied exclusively with EU-27 countries: Greece, Cyprus, Spain, Portugal and Poland. These countries may be presumed to be importers of foreign science, but – as announced – more insight into cross-country flows is provided in a next analytical section.

Countries in the lower left area have a modest national science base, and a corporate texture which sources relatively little knowledge from scientific sources. BRIC countries are situated here, along with South Africa and Asian countries (Taiwan, Korea and Japan). It is noteworthy that technologically strong countries like Japan and Italy are represented in this quadrant as well.

Further diagnosis of countries' positions in the different quadrants requires an insight in cross-country flows of science towards corporate technology. This is the focus in the following analytical section.

### 3.2. Mapping of cross-country citation flows between science and technology (based on patents citing scientific literature)

Whereas the previous section addressed national-level relations between supply and absorption of science, it made abstraction of the question *where* the cited science is sourced. This section sheds light on the geographic dimension of technology citing science by focusing on cross-country citation flows. The main questions relate to the presence of a home bias or of geographic proximity effects in the citations from a country's corporate patents towards scientific resources.

The deployed indicators and analyses are derived from the results of the above-described matching between patents and scientific documents. On the citing side (patents), EPO, USPTO and WIPO corporate patents are considered with application years between 2000 and 2009. Restricting the analyses to corporate patents allows for a clear focus on citation flows towards the industrial texture. On the cited side (scientific articles), it is important to note that our available Web of Science data infrastructure is restricted to publications after 1990. Therefore, NPRs can only be matched to scientific publications from 1991 onwards. To give an idea of underlying volumes, table 4 reports the set of WoS references that is matched against, i.e. the yearly worldwide volumes of scientific publications in the Web of Science without excluding any publication type (article, review, meeting abstract, note, letter, ...) or discipline (Sciences, Social Sciences, Arts & Humanities).

Volume Year	# WoS documents	Volume Year2	# WoS documents2
1991	897.264	2002	1.170.168
1992	947.102	2003	1.294.646
1993	949.654	2004	1.266.310
1994	1.004.538	2005	1.553.041
1995	1.069.736	2006	1.446.205
1996	1.128.646	2007	1.492.276
1997	1.148.847	2008	1.687.758
1998	1.171.805	2009	1.726.030
1999	1.179.260	2010	1.637.900
2000	1.165.682	2011	1.760.015
2001	1.204.955		

**Table 4: yearly volumes of scientific publications in the Web of Science (1991-2011)**

For mapping cross-country citation linkage patterns, citation matrices were developed that map countries of citing corporate patents (rows) and countries of origin of the cited scientific references (columns). Global coverage is assured by including OECD member countries, EU-27 (and EFTA) member states, EU candidate

countries, BRIC countries and Taiwan (on citing and cited side)<sup>6</sup>. Table 5 shows cross-country citation linkages between these countries, whereby the cell values represent relative intensities of citation linkages between the citing countries (rows) and the cited countries (columns). By using relative intensities rather than absolute numbers of citation links<sup>7</sup>, size effects are corrected for, allowing for adequate inter-country comparison of citation patterns. The following formula was used to calculate these relative intensities:

$$CCC_{ij} = \frac{\frac{\text{\#citations of corporate citing country i to science cited country j}}{\text{total \#citations of corporate citing country i}}}{\frac{\text{citations to science cited country j}}{\text{total corporate citations}}}$$

A value higher than 1 represents a link between citing and cited country which is “overrepresented”. The framed figures on the diagonal are of particular interest, as they indicate within-country citation flows which will be used to evaluate a home bias in citation patterns.

---

<sup>6</sup> This results in a dataset comprising 51 countries: Australia, Austria, Belgium, Brazil, Bulgaria, Canada, Chile, Croatia, Chile, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, India, Ireland, Israel, Italy, Japan, Latvia, Liechtenstein, Lithuania, Luxembourg, Macedonia, Malta, Mexico, Montenegro, the Netherlands, New Zealand, Norway, China, Poland, Portugal, Romania, Russia, Serbia, Slovakia, Slovenia, South Africa, South Korea, Spain, Sweden, Switzerland, Taiwan, Turkey and the United Kingdom. At the same time, after data gathering, several countries were eliminated from the analyses, either because their volumes of citations given or received are below 100, or because their population is smaller than 300.000 persons. These low values cause relative / normalized indicators for these countries to be artificially high, thereby distorting the analyses.

<sup>7</sup> For the matrix with cell values representing the absolute numbers of cited-citing patent-article pairs for the respective country combinations, we refer to appendix 3.



A visual inspection of the matrix reveals that, overall, countries cite their own science more intensively than science from foreign countries (cell values on the diagonal representing higher intensities). At the same time, it becomes apparent that the extent of such a 'home bias' differs between countries. More specifically, countries where the within-country citation intensity is lowest are US, UK, DE (index below 2) and JP, FR, CH, CA, NL (index below 3). Most of these countries are at the same time technologically strong, suggesting that reliance on foreign science leverages technological performance. A more thorough analysis of the implied premise is conducted in section 3.3 (relation between characteristics of ST linkage patterns and technological performance). When considering citation linkages beyond country borders, the matrix suggests that geographic proximity seems to matter. Examples of groups of neighboring countries that display relatively strong citation linkage intensities include: Austria, Switzerland, Germany – Denmark, Norway, Sweden – Sweden, Finland – France, Belgium.

To statistically verify the above observations, an analysis of covariance (ANCOVA) was performed. We recall that one of the focal questions was whether knowledge sourcing in national technology development is characterized by a home bias in the citations towards science. The dependent variable is the index of relative citation intensity between citing and cited country. It is transformed logarithmically to meet distributional requirements for the model. Independent variables are: home bias; geographic distance between citing and cited country and difference in scientific productivity between citing and cited country. The home bias variable is represented by a dummy for 'within-country' (1) versus 'foreign' (0) citation links. If a home bias is present, a positive relation between 'home' and citation intensity is expected. Based on the observation in the matrix (table 5) that the home bias appears more outspoken for some countries, an additional interaction effect was introduced between citing country and home bias. Turning attention to cross-border citation patterns, a geographic distance measure is included to assess the strength of a proximity effect in technology-science citation patterns. The measure has been adopted from the CEPII<sup>8</sup> GeoDist data files (see <http://www.cepii.fr/anglaisgraph/bdd/distances.htm>). These files contain measures of bilateral distances (in kms) for most countries of the world. The measure introduced in our model is a weighted distance measure (transformed logarithmically to meet distributional requirements for the model). It calculates distances between two countries based on bilateral distances between their biggest cities, whereby the inter-city distances are weighted by the share of the city in the overall country's population. If a geographic proximity effect is present in science-technology citation linkages, then a negative relation between this distance measure and citation linkage intensity is expected. The model includes controls for country specificities (citing and cited country), and domain-specificities are

---

<sup>8</sup> French research center in international economics; See: [www.cepii.fr](http://www.cepii.fr)

accounted for by introducing technological domains, classified into 35 classes according to the classification developed by ISI-Fraunhofer, Schmoch, 2008: see Appendix 2). Results are presented in table 6.

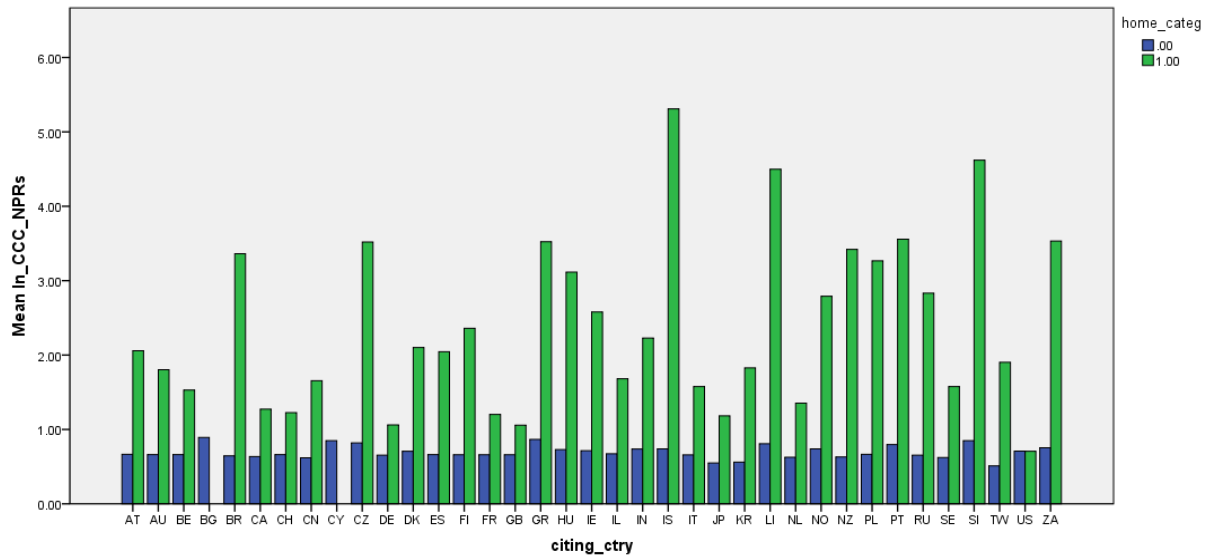
<b>ANCOVA - Tests of Between-Subjects Effects</b>						
<b><i>Dependent Variable: Relative citation intensity between citing and cited country</i></b>						
<b>Source</b>	<b>Type III Sum of Squares</b>	<b>df</b>	<b>Mean Square</b>	<b>F</b>	<b>Sig.</b>	<b>Parameter Estimate (B)</b>
Corrected Model	3004,044 <sup>a</sup>	136	22,089	116,576	,000	
Intercept	301,668	1	301,668	1592,107	,000	2,738
Home (1/0)	358,625	1	358,625	1892,709	,000	1,530
Geographic distance between citing and cited country	17,118	1	17,118	90,343	,000	-,046
Citing country * Home	217,972	32	6,812	35,950	,000	
Technology Domain (FhG35)	614,326	34	18,068	95,359	,000	
Citing Country	840,003	34	24,706	130,390	,000	
Cited Country	556,623	34	16,371	86,402	,000	
Error	3219,408	16991	,189			
Total	19716,133	17128				
Corrected Total	6223,452	17127				

a. R Squared = ,483 (Adjusted R Squared = ,479)

**Table 6: Ancova analysis: presence of home bias in citation linkage patterns between corporate patents and scientific references.**

The results confirm the presence of a home effect in science-technology linkage patterns. Technological development in a country indeed seems to link primarily to the own national science base, a finding which supports the existence of the national innovation 'systems' notion. At the same time, the significant interaction effect with citing country confirms that there are country differences in the extent of the home bias. In addition, the negative effect of the geographic distance measure shows that proximity matters within foreign science-technology linkages. This was already suggested in the above matrix (table 5), where some clusters of higher citation intensities between neighboring countries became apparent (cf. supra).

As said, the significant interaction effect between home bias and citing country confirms that the home bias is not equally outspoken for all countries. This is further illustrated in Figure 3, which compares the within-country citation intensity to the average citation intensity with foreign countries, for each citing country.



**Figure 3: Comparison of within-country (home) citation intensity (green) to the average citation intensity with foreign countries (blue),**

Figure 3 shows that, although the overall tendency clearly shows higher within-country citation intensities, this difference between home and foreign is less outspoken for countries like US, France, UK, Japan and Germany. Most of these countries are at the same time technologically strong, suggesting that reliance on foreign science leverages technological performance. A more thorough analysis of the implied premise is conducted in section 3.3 (relation between characteristics of ST linkage patterns and technological performance).

Finally, the results in table 6 reveal significant technology domain specificities in science-technology linkage patterns. To illustrate these, the cross country citation matrix (see table 5) was broken down into technological domain (35 Fraunhofer classes<sup>9</sup>). These domain-specific matrices are added in annex 1a. In addition, annex 1b provides a breakdown of figure 3 (home bias by country) per Fraunhofer domain. To summarize observations on technological specificity of the home bias, Table7 summarizes within-country citation intensities for each citing country, by technology field. It can be seen that also within countries, there is considerable variety in the strength of a home bias in scientific prior art, depending on the technological domain. For example, Germany has an outspoken home bias in Civil Engineering, China in Food Chemistry, the Netherlands in Mechanical Elements; UK and France in Machine tools. For the US, the low home bias that was observed earlier on appears to be relatively domain-independent. Further diagnosis of the domain-specificity in these linkage patterns and of cross-border science-technology linkages, would benefit from a breakdown in domains not only at the technology (patent) side, but also at the science side. A weaker home bias may reflect access to broader international networks in complement of a country's own science base,

<sup>9</sup> Technology classification developed by ISI-Fraunhofer, Schmoch, 2008 (see Appendix 2)



but it may also reflect a mismatch between the developed technologies and the science base present within the home country. Such a diagnosis becomes extra relevant in light of the current research and policy initiatives evolving around smart specialization. At the same time, analyses of that kind require the use of concordance schemes mapping relations between technological and scientific domains. The results of the NPR-WoS matching on which this report is based, are currently being used for developing such concordance schemes (by identifying which scientific disciplines are cited by which technological domains). Hence, although it is beyond the scope of the current analytical report, further diagnosis of domain specificity in citation linkages will become possible in the near future, after validation and consolidation of the resulting concordance.

	AT	AU	BE	BR	CA	CH	CN	CZ	DE	DK	ES	FI	FR	GB	GR	HU	IE	IL	IN	IS	IT	JP	KR	NL	NO	NZ	PL	PT	RU	SE	SI	TW	US	ZA	
1		7,95	5,41	2,80		3,58	3,37	3,26		2,01	12,68	6,30	15,87	2,05	2,61	92,82	117,99	16,11	10,63	11,91		2,37	1,78	2,88	4,54		44,86	19,99		2,86	5,41		6,32	1,05	78,44
2		37,43	2,69	3,88		2,45	2,57			2,36	4,54		5,73	1,48	2,65			5,57	1,60			5,45	1,67	3,32	2,53	7,04				6,90		3,66	1,04		
3			0,27	11,66		1,63	2,64	2,47		1,78	6,57	11,08	4,43	2,12	1,74			10,07	1,21			3,24	1,81	2,72	2,71	11,19				3,90		2,10	1,02		
4		6,48	1,43	18,39		1,34	1,89	2,31		1,91	6,55	11,08	2,74	2,23	2,27			12,46	0,30			2,05	1,92	2,87	2,42	19,57				2,63		1,72	1,04		
5		60,33	1,60	2,82		1,89	1,98	2,42		1,44			2,79	2,94	2,21				3,47	48,27		3,83	1,92	1,83	2,95		423,80		560,52	4,57		2,42	1,04		
6		6,73	3,72	5,20		2,29	2,27	2,32		1,73	10,10	8,29	7,11	2,63	1,72			17,68	1,82	2,60	240,25	2,92	1,90	3,39	2,95	7,92	26,99	40,94		4,89		1,38	1,02		
7			2,19			0,86	0,90			1,76	10,92	13,71	10,75	2,40	2,07				4,73			7,33	1,62	3,25	2,37	298,84							1,01		
8		10,97	24,08	2,81		4,17	5,08	1,35		1,65	30,43	7,03	10,47	2,86	2,58				2,19			3,07	1,74	2,53	2,24		112,06	34,82		6,43		2,24	0,99		
9		19,90	14,79	1,65		3,34	4,79	3,08		1,92	12,55	27,04	26,13	2,79	2,15		38,15	22,98	3,85	27,85		4,69	1,83	3,55	2,60		34,37	42,48		8,90		5,25	1,01		
10		8,76	6,90	7,25	58,64	2,90	3,13	4,72		2,03	6,38	7,18	15,54	2,91	2,23	118,86	40,65	23,74	4,45	31,12	378,64	4,47	2,43	6,39	2,34	12,07	15,68	13,10	30,31	15,13	4,63	11,34	1,01		
11		5,80	4,24	4,34	43,11	2,36	1,76	6,52	30,31	1,66	5,59	6,25	11,38	2,42	1,81	33,07	33,13	7,05	7,53	32,49	156,32	4,04	2,35	10,36	3,11	13,71	31,25		21,70	24,48	3,71	162,87	15,25	0,99	
12				7,44		2,30	2,59			1,90		8,39		2,97	2,34				6,47	28,59		1,98	1,93	2,32	2,89					11,80		4,89	1,04		
13		5,67	4,50	2,82		2,54	1,15	6,07		1,96	7,89	10,00	13,21	2,68	2,11			8,21	2,53	4,97		3,67	2,58	12,99	1,99	14,32	51,32	25,49		4,52		4,02	0,98		
14		6,57	4,71	3,00	7,24	2,81	2,15	3,93		1,86	6,85	5,00	14,75	2,08	1,73	122,48	9,54	7,88	3,68	3,88	125,60	3,08	2,10	6,44	2,47	15,42	31,47	5,85	51,48	12,71	4,03	21,83	9,08	0,99	
15		5,90	3,81	4,11	23,73	2,75	1,98	4,72	26,01	1,76	4,23	5,94	10,52	2,35	1,84	11,42	20,90	12,06	4,06	13,77	129,61	4,67	2,41	6,46	3,15	10,61	14,42	16,13	23,13	21,72	3,66	229,77	10,83	0,98	
16		4,89	4,66	2,84	44,90	2,41	2,16	6,45	42,65	1,85	5,56	5,36	15,54	2,32	1,65	24,24	14,69	9,87	5,86	8,00	147,22	3,56	2,42	9,20	3,13	13,86	26,95	21,97	36,52	20,64	2,89	137,19	16,59	1,00	
17		3,29	5,79	2,09	25,41	3,16	3,04	4,41	116,45	1,66	9,34	10,85	6,50	2,74	2,31	71,27			6,12	7,27	329,94	5,70	1,86	4,56	3,20	13,17	42,42				9,88		23,36	1,02	
18		11,83	3,29	3,86	11,75	2,92	2,34	10,77		2,05	3,93	6,19	12,75	2,10	1,48	290,80	18,63	18,92	1,78	4,97		4,14	3,02	9,37	2,57	12,23	8,01		42,41	3,41		21,50	1,08		
19		9,14	5,18	1,70	33,41	2,37	2,78	4,53		1,84	6,18	4,79	11,81	2,19	1,60			93,65	15,95	7,02	5,60	124,66	4,22	2,08	5,11	2,27	15,66	24,01			3,62	141,87	8,21	1,05	
20		10,46	9,91	1,52	60,86	3,47	3,94	7,71		2,25	48,03	2,55	29,99	2,01	1,95				11,71	6,07		3,21	1,84	3,79	1,51	6,80	55,43	21,37		10,42	11,13		2,66		
21		6,02	7,55		81,35	3,47	2,61	8,50		2,18	41,13	10,72	13,02	2,38	2,68	274,55			10,47			4,34	1,98	3,29	2,69	30,81	34,86	18,18		11,00	7,12		2,30		
22			5,61			3,14	4,70			1,55	23,39	20,77	14,24	2,76	1,49				3,16			1,58	2,19	4,10	1,59		27,29			17,99	7,84		7,00		
23		9,07	9,03	6,18	31,20	3,03	2,42	2,73	62,55	1,93	13,31	12,79	18,66	2,41	1,84		31,06	6,40	5,51	7,61		2,49	2,40	6,88	2,39	14,93	69,07	29,47	42,10	19,36	7,73	94,72	12,98		
24		10,90	6,64		56,37	1,56	0,80	8,51		2,34	6,10	13,15	12,52	2,59	1,84				4,44	18,79		6,66	2,28	5,35	2,00	27,52		15,98		12,65	7,69	385,22	17,69		
25			5,16			1,31	0,84			1,47		15,53		2,38	1,32		245,83					5,42	1,93		2,79	17,35					5,36		1,08		
26		9,05	4,93	12,13		4,07	1,90	7,92		1,79	18,19			4,78	4,61						7,04	1,69	4,91		295,58			107,48			9,84		1,07		
27		20,09	6,13	4,47		1,84	4,01			2,06	9,65	5,41	12,68	3,01	1,83				3,94	36,78		0,81	2,47	9,94	2,29	31,05				7,44		21,35	1,02		
28				2,16		5,53	4,76	7,50		1,70	10,95	9,21	9,88	1,66	2,71			14,52	14,97		7,20	1,90	2,72	1,89	12,28	19,35				10,64		0,94			
29		2,82	3,64	2,82		2,50	3,07	2,57	123,30	2,02	7,58	4,22	18,19	2,29	1,80	95,55	18,97	8,33	20,52	339,08	6,77	2,10	6,49	3,85	14,95	22,21			18,50	4,18		6,24	1,05		
30		29,11				7,18	6,48			2,13			11,06	3,40	1,88							2,06		4,15							28,38		1,02		
31						4,18				1,95	33,29			3,91	0,90				53,27			7,36	1,59	20,15	15,77	28,20					6,66		1,06		
32		25,45				4,49	8,55			1,51		13,57		1,81	1,83				40,72			0,96	1,44		2,19					28,28		1,13			
33			3,34			5,48				0,55			47,57		0,94							1,31		6,17									0,94		
34						2,21	3,91	12,88		1,74			13,97	4,43	1,41	235,70		4,59				4,85	2,58	3,57	0,87					6,03		0,96			
35						1,67				3,84			3,10	1,47								5,05		4,76	24,26					11,23		1,03			

Table 7: overview of within country science based CCC-values per FgH35 technology field

To summarize, the results from this section clearly confirm the presence of a home bias and of a (be it weaker) cross-border proximity effect in science cited in patents. The presence of a home bias varies between countries and technological domains. For several countries that are traditionally known as technologically strong (e.g. US, DE,...), the home bias is less outspoken.

The presence of a home bias implies that countries can rely on their own scientific knowledge base for advancing their technological development and innovative activities, hence providing support for the existence of the national innovation 'systems' notion. At the same time, the home bias is not absolute. The findings also reveal that science-technology citation linkages cross national borders and that they often occur on a global scale.

Especially relevant for policymakers is the extent to which the presence of a home / proximity bias and characteristics of foreign science-technology linkage patterns affect national technological performance. This is the subject of the third analytical section.

### **3.3. Analysis of the relation between national technological performance and science-technology linkage patterns on a national level: relations between ST linkage patterns and national patent output.**

In this section, the relationship between characteristics of science-technology patterns and national technological performance are analyzed. National technological performance on the one hand is measured by a country's patent volume per capita. For measuring characteristics of ST linkage patterns, a number of national-level variables have been developed, based on the analyses reported in the previous sections.

A first characteristic concerns the strength of a country's 'home bias' in citing science, i.e. the extent to which a country's patents contain references to its own scientific articles versus references to scientific articles from foreign countries. The measure used is the ratio between 'within country citation intensity' (CCC\_home, see the green bars in Figure 3) and the average citation intensity between the source country and cited foreign countries (CCC\_foreign, represented by the blue bars in Figure 3).

Second, as a complement to the home bias variable, a Herfindahl-based measure was developed that represents the concentration of foreign citations over cited countries. The Herfindahl index is the sum of the squares of the (received) citations shares of all cited foreign countries, where the citation shares are expressed as fractions of the total number of citations to foreign countries. The result is proportional to the average citation share, weighted by citation share. As such, it can range from 0 to 1, moving from a large number of foreign countries each receiving a small number of citations (maximum geographic breadth), to a single foreign country receiving all citations (maximum concentration). In other words, a higher Herfindahl index indicates that science is cited from a smaller, more concentrated set of countries (the more geographically focused foreign citing patterns). A lower Herfindahl index indicates a geographically broader pattern of foreign citations.

Several control variables are added to the model. By introducing technological domains (according to the Fraunhofer classification in 35 domains), we control for domain-specific effects. Moreover, several characteristics of national science and technology textures are considered that may be related to national technological performance. 'Scientific Output per 1000 Capita' represents the country's productivity in scientific research, measured by the number of scientific articles in the Web of Science for the period 1991-2009 (cf. supra), and normalized for

country size. 'Number of SNPRs' represents the number of scientific citations within a country's patent portfolio, broken down by technological domain (EPO, US, PCT patents, 2000-2009). Finally, 'national number of university patents' is another indicator of national science-technology relatedness, representing the extent to which universities contribute to national technology development (EPO, US, PCT patents involving at least one university applicant<sup>10</sup>, 2000-2009). The results of the ANCOVA analysis, along with estimates of the parameters (B), are presented in table 8.

<b>ANCOVA - Tests of Between-Subjects Effects<sup>11</sup></b>						
<b>Dependent Variable: Number of Patents per mio capita (Citing Country)</b>						
<b>Source</b>	<b>Type III Sum of Squares</b>	<b>df</b>	<b>Mean Square</b>	<b>F</b>	<b>Sig.</b>	<b>Parameter Estimate (B)</b>
Corrected Model	2282,367 <sup>a</sup>	39	58,522	100,608	,000	
Intercept	2,230	1	2,230	3,834	,050	1,221
Scientific Output per 1000 Capita (Citing Country)	656,199	1	656,199	1128,096	,000	2,958
Number of university patents per mio capita (Citing Country)	20,438	1	20,438	35,136	,000	,245
Number of SNPRs (Citing Country)	79,180	1	79,180	136,121	,000	,218
Herfindahl Foreign Citations	2,710	1	2,710	4,658	,031	-,435
Ratio Home versus Foreign citation intensity	,091	1	,091	,156	,693	-,013
Technology domain (FhG35)	298,384	34	8,776	15,087	,000	
Error	566,563	974	,582			
Total	19157,226	1014				
Corrected Total	2848,930	1013				
a. R Squared = ,801 (Adjusted R Squared = ,793)						

**Table 8: Ancova – Relation between national-level characteristics of ST linkage patterns and national technological performance**

<sup>10</sup> Identification of university patents was based on the sector allocation methodology, developed at ECOOM – KU Leuven (see: Eurostat (2011), 'Patent Statistics at Eurostat: Methods for Regionalisation, Sector Allocation and Name Harmonisation', Methodologies & Working papers, Publications Office of the European Union, Luxembourg, 2011, ISBN 978-92-79-20237-7).

<sup>11</sup> All continuous variables have been transformed logarithmically for better fitting distributional characteristics. For patents with multiple applicants from different countries, a full-fractional counting scheme is used (whereby each represented country is counted as 1). For patents belonging to more than 1 Fraunhofer domain, a full counting scheme is used.

Controlling for significant effects of technology domains, the results confirm that several characteristics of national science-technology systems are related to national technological performance. The relation between national scientific productivity and national technological performance is most outspoken. As for measures of national science-technology relatedness, the number of university patents is positively related to national technological output. In addition, the volume of SNPRs present in the national technology portfolio, as a second measure of national-level science-technology relatedness, shows a significant positive relation with technological performance. The positive relation between national-level measures of science-technology interaction on the one hand and national technological performance on the other hand confirms previous findings (see e.g. Van Looy et al., 2006).

Turning our attention to the indicators of cross-country linkage characteristics, the results show no significant relation between the extent of the home bias and national technological performance. The results for the Herfindahl index suggest the appropriate modality of foreign citation patterns. The Herfindahl index for geographic concentration in foreign citations is negative and significant. This means that countries of which the cited science base is spread over a broader set of (foreign) countries are the technologically more performing countries. In other words, a sufficient degree of geographic diversity in cited foreign scientific prior art appears to be more beneficial than a science base that is concentrated in fewer foreign countries, in terms of leveraging national technological performance.

The lack of a relation between home bias and national technological performance may seem surprising, when considering the observations from table 7 and figure 3. These observations suggested that a lesser extent of home bias is related to higher technological performance, which is not confirmed by the model in table 8. At the same time, the effects of ST linkage pattern characteristics (extent of home bias and Herfindahl concentration index of foreign citations) may be interrelated with the scientific capacity from the home country. One could e.g. assume that scientifically weaker countries require stronger reliance on foreign science to keep their technology development up to pace. In order to test this assumption, interaction effects between the extent of the home bias and the Herfindahl index on the one hand and scientific productivity of the citing country on the other hand were additionally introduced into the model. The results are presented in table 9.

<b>ANCOVA - Tests of Between-Subjects Effects</b>						
<b>Dependent Variable: Number of Patents per mio capita (Citing Country)</b>						
<b>Source</b>	<b>Type III Sum of Squares</b>	<b>df</b>	<b>Mean Square</b>	<b>F</b>	<b>Sig.</b>	<b>Parameter Estimate (B)</b>
Corrected Model	2290,597 <sup>a</sup>	41	55,868	97,261	,000	
Intercept	7,185	1	7,185	12,509	,000	1,544
Scientific Output per 1000 Capita (Citing Country)	98,794	1	98,794	171,991	,000	2,362
Number of university patents per mio capita (Citing Country)	22,371	1	22,371	38,945	,000	,259
Number of SNPRs (Citing Country)	83,244	1	83,244	144,919	,000	,226
Ratio Home versus Foreign citation intensity	4,039	1	4,039	7,032	,008	-,156
Herfindahl Foreign Citations	8,354	1	8,354	14,544	,000	-1,330
Technology domain (FhG35)	299,064	34	8,796	15,313	,000	
Scientific Output per 1000 Capita (Citing Country)* Herfindahl Foreign Citations	5,744	1	5,744	10,000	,002	1,528
Scientific Output per 1000 Capita (Citing Country)* Ratio Home versus Foreign citation intensity	4,661	1	4,661	8,114	,004	,231
Error	558,333	972	,574			
Total	19157,226	1014				
Corrected Total	2848,930	1013				
a. R Squared = ,804 (Adjusted R Squared = ,796)						

**Table 9: Ancova – Relation between national-level characteristics of ST linkage patterns and national technological performance, including interaction effects**

When introducing interaction effects with scientific productivity of the citing country, a main effect of the home bias indeed becomes significant. The main effect is negative, signaling that – overall – more reliance on foreign science is beneficial for national technology development activities. At the same time, the interaction effect is significant, implying that the relation between home bias and national technological productivity depends on the level of scientific capabilities of the citing country. The main effect of the Herfindahl index of geographic concentration in foreign citations remains negative and becomes stronger, again confirming that a broader geographic scope in citing science is beneficial for national technological activities. But also here, the interaction effect between the Herfindahl index and scientific capacity of the citing country is significant. Hence, the relation between the Herfindahl index of geographic concentration in foreign citations on the one

hand and national technological productivity on the other hand depends on the level of scientific capabilities of the citing country.

For a transparent interpretation of the interaction effects, the analyses were split up in two groups: countries in the lower 50% percentile of scientific productivity versus countries in the upper 50% percentile. The results are presented in table 10. It becomes clear that the positive effect of foreign scientific citations and a broad geographic scope in the origin of these foreign citations on national technological performance is valid for countries with a smaller own science base. For countries with large scientific capabilities, these leveraging effects of foreign and broad science citation patterns are not decisive for national technological performance.



ANCOVA - Tests of Between-Subjects Effects												
<i>Dependent Variable: Number of Patents per mio capita (Citing Country)</i>												
	LOW SCIENTIFIC PRODUCTIVITY						HIGH SCIENTIFIC PRODUCTIVITY					
Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Parameter Estimate (B)	Type III Sum of Squares	df	Mean Square	F	Sig.	Parameter Estimate (B)
Corrected Model	1184,370 <sup>a</sup>	39	30,368	51,652	,000		283,693 <sup>a</sup>	39	7,274	22,809	,000	
Intercept	,500	1	,500	,850	,357	1,132	7,146	1	7,146	22,408	,000	1,971
Scientific Output per 1000 Capita (Citing Country)	234,845	1	234,845	399,438	,000	3,702	53,635	1	53,635	168,178	,000	2,645
Number of university patents per mio capita (Citing Country)	45,655	1	45,655	77,653	,000	,773	,384	1	,384	1,204	,273	,043
Number of SNPRs (Citing Country)	20,956	1	20,956	35,644	,000	,180	14,160	1	14,160	44,399	,000	,135
Ratio Home versus Foreign citation intensity	3,277	1	3,277	5,574	,019	-,102	,001	1	,001	,003	,958	,002
Herfindahl Foreign Citations	3,327	1	3,327	5,659	,018	-,619	,074	1	,074	,232	,630	-,134
Technology domain (FhG35)	149,561	34	4,399	7,482	,000		159,305	34	4,685	14,692	,000	
Error	267,513	455	,588				147,978	464	,319			
Total	6009,784	495					12961,258	504				
Corrected Total	1451,882	494					431,671	503				
	a. R Squared = ,816 (Adjusted R Squared = ,800)						a. R Squared = ,657 (Adjusted R Squared = ,628)					

**Table 10: Ancova – Relation between national-level characteristics of ST linkage patterns and national technological performance, split by level of scientific performance (Low versus High)**

To summarize, technologically strong countries are characterized by high levels of scientific productivity and science-technology relatedness (measured by university patenting and scientific prior art in patent documents). Indeed, effects are largely driven by the scale of a country's scientific and technological activities. In addition, it is shown that characteristics of cross-country linkage patterns are related to technological performance on a national level. The presence of foreign linkage patterns to complement home-based knowledge sourcing, as well as a broad geographical scope in citing foreign science are positively related to technological performance. However, a significant interaction effect reveals that leveraging effects of this foreign orientation in citing science are primarily relevant for scientifically weaker countries. Countries that have a large own science base at their disposal, appear much less dependent on foreign sourcing and a broad geographic scope for enhancing or maintaining their technological performance.

## 4. SUMMARY CONCLUSIONS & DIRECTIONS FOR FOLLOW-UP STUDIES

This analytical report presents the first findings obtained from a large scale matching effort between scientific references cited as prior art in patent documents on the one hand and individual articles in the Web of Science on the other hand. The developed matching algorithm succeeds in uniquely matching almost 70% of non-patent references for which sufficient information is available ('matchable NPRs'). In absolute figures, for patents with an application year between 1993 and 2009, this comes down to almost 3.3 million non-patent references for which a matched article was identified in the Web of Science. Such a result is – according to the authors' knowledge – unprecedented and entails great potential for large-scale analysis of science-technology linkage patterns on several levels of analysis. The current report considers cross-country linkages (coverage of major world regions) between citing technology (patents) and cited science (articles), which is highly relevant for addressing questions on the sources of national technological development as drivers behind a nation's innovative performance. In what follows, we summarize the findings relating to three main research questions.

1. Do supply and absorption of scientific resources for technology development coincide on a national level?

The analyses on the supply and absorption of scientific resources for technology development revealed that, although there is a slight positive relation between both, several countries are characterized by discrepancies between the extent of their own science base and the degree to which technology development in their corporate texture is related to scientific research (measured by scientific citations in corporate patents). Most European countries display profiles that combine a high scientific productivity, combined with a considerable absorption of science in national technology development activities. These figures may hence be indicative of many European countries representing efficient and dynamic innovation systems that seek to exploit knowledge through close science-technology interactions. Moreover, except for some Scandinavian countries with slightly lower absorption rates of science in corporate technology, none of the individual countries are revealed to be national-level cases of the European Paradox (severe discrepancy between large science supply and minimal absorption by the corporate texture).

2. What are patterns of cross-country citation flows between science and patents?

This section considered cross-country flows between citing corporate patents and cited science, by mapping relative citation intensity indices for pairs of citing and cited countries. The results clearly confirmed the presence of a significant home

bias and of a (be it weaker) cross-border proximity effect in science cited in patents. The presence of a home bias varies between countries and technological domains. For several countries that are traditionally known as technologically strong (DE, FR, UK,...) and especially for the US, the home bias is clearly less outspoken.

The presence of a home bias implies that countries can rely on their own scientific knowledge base for advancing their technological development and innovative activities, hence providing support for the existence of the national innovation 'systems' notion. At the same time, the home bias is not absolute. The findings revealed that science-technology citation linkages do cross national borders and, even more, that they often occur on a global scale. These findings are in line with a previous study, conducted by Veugelers et al. (2012) in which cross-country citation patterns between corporate (citing) patents and academic (cited) patents were analyzed.

### 3. What are the relations between national-level characteristics of science-technology linkage patterns and national-level technological performance?

The results confirm that characteristics of cross-country linkage patterns are related to technological performance on a national level. The presence of foreign linkage patterns to complement home-based knowledge sourcing, as well as a broad geographical scope in citing foreign science are positively related to technological performance. However, these leveraging effects of foreign orientation in citing science are primarily relevant for scientifically weaker countries.

To conclude: for enhancing and maintaining technological performance, sufficient levels of linkage to science appear imperative. The modalities of this linkage between technology and science matter, whereby a broader geographical scope in scientific prior art is beneficial for leveraging technological performance.

The present findings are based on cross-sectional data, comparing countries for an aggregated period of 10 years. This impedes more sophisticated conclusions on causal relations between the variables. E.g. does a national-level increase in scientific activities lower the extent of linkage to foreign science? Are higher levels of absorption of science in a country's technological activities the result of an increased supply in science? For being able to answer such questions and get a more dynamic perspective on ST linkage patterns, follow-up studies are in place, based on longitudinal data. Panel data allow for fixed effects models, to control for unobserved country heterogeneity. Moreover, they allow for introducing time lags between dependent and independent variables, which is a way for answering questions on causality between variables.

## REFERENCES

Adams, J. 1990. Fundamental stocks of knowledge and productivity growth. *Journal of Political Economy* 98, 673–702.

Alcacer & Gittelman, 2006 Patent citations as a measure of knowledge flows: The influence of examiner citations. *Review Of Economics And Statistics* Vol. 88 Issue: 4 Pages: 774-779

Audretsch, D. B. and P. E. Stephan. 1996. Company-scientist locational links: The case of biotechnology. *The American Economic Review* 86, No 3: 641–652.

Baumol, W.J. 2002. *The Free-Market Innovation Machine: Analyzing the Growth Miracle of Capitalism*. Princeton University Press.

Breschi & Lissoni, 2001, Localised knowledge spillovers vs. innovative milieux: Knowledge "tacitness" reconsidered, *Papers In Regional Science* Volume: 80 Issue: 3 Pages: 255-273

Callaert, J., Grouwels, J., & Van Looy, B. (2011), "Delineating the scientific footprint in technology: Identifying science within non-patent references," *Scientometrics*, 91 (2), 383-398.

Callaert, J., Van Looy, B., Verbeek, A., Debackere, K., Thijs, B. (2006). "Traces of Prior Art: An analysis of non-patent references found in patent documents." *Scientometrics*, 69 (1), 3-20.

Du Plessis, M., Van Looy B., Song X. Magerman, T. (2009). *Data Production Methods for Harmonized Patent Statistics: Patentee Sector Allocation*. Eurostat Working Paper and Studies, Luxembourg. [https://lirias.kuleuven.be/bitstream/123456789/228565/1/msi\\_0606.pdf](https://lirias.kuleuven.be/bitstream/123456789/228565/1/msi_0606.pdf)

Eurostat, 'Patent Statistics at Eurostat: Methods for Regionalisation, Sector Allocation and Name Harmonisation', *Methodologies & Working papers*, Publications Office of the European Union, Luxembourg, 2011, ISBN 978-92-79-20237-7.

Freeman, C. 1987. *Technology and Economic Performance: Lessons from Japan*. Pinter, London.

Freeman, C. 1994. The economics of technical change: A critical survey article. *Cambridge Journal of Economics* 18, No.5: 463–514.

Jaffe, A.B., Trajtenberg, M., & Henderson, R. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics* 108, No. 3: 577-598.

Lundvall, B.-Å. 1992. *National Systems of Innovation*. Pinter, London.

Meyer, 2000, Patent citations in a novel field of technology - What can they tell about interactions between emerging communities of science and technology? *Scientometrics*, Volume: 48 Issue: 2 Pages: 151-178.

Mowery, D.C., & Nelson, R.R. 1999. *Sources of Industrial Leadership*. Cambridge University Press, Cambridge.

Nelson, R.R. 1993. *National Innovation Systems*. Oxford University Press, New York.

Nelson, 2009, Measuring knowledge spillovers: What patents, licenses and publications reveal about innovation diffusion. *Research Policy*, Volume: 38 Issue: 6 Pages: 994-1005.

OECD Patent Statistics Manual, February 2009

<http://browse.oecdbookshop.org/oecd/pdfs/free/9209021e.pdf>

Schmoch, U., Concept of a technology classification for country comparisons, Final report to the World Intellectual Property Organization, Fraunhofer Institute for Systems and Innovation Research, June 2008  
[http://www.wipo.int/edocs/mdocs/classifications/en/ipc\\_ce\\_41/ipc\\_ce\\_41\\_5-annex1.pdf](http://www.wipo.int/edocs/mdocs/classifications/en/ipc_ce_41/ipc_ce_41_5-annex1.pdf)

Tijssen RJW; Buter, RK; van Leeuwen, TN. 2000, Technological relevance of science: An assessment of citations linkages between patents and research papers. *Scientometrics* , Volume: 47 Issue: 2 Pages: 389-412.

Van Looy, B., Callaert, J., Debackere, K., & Verbeek, A. (2002). "Patent-related indicators for assessing knowledge-generating institutions: Towards a contextualised approach," *The Journal of Technology Transfer*, vol. 28, no. 1, pp. 53-61.

Van Looy, B., Debackere, K., Callaert, J., Tijssen, R., van Leeuwen, T. (2006), "National scientific capabilities and technological performance: An exploration of emerging industrial relevant research domains." *Scientometrics*, 66 (2), 295-310.

Van Looy Bart; Magerman, Tom; Debackere, Koenraad. 2007, Developing technology in the vicinity of science: An examination of the relationship between science intensity (of patents) and technological productivity within the field of biotechnology. *Scientometrics*, Volume: 70 Issue: 2 Pages: 441-458.

Verbeek, A., Debackere, K., Luwel, M., Andries, P., Zimmermann, E., & Deleus, F. (2002). Linking science to technology: Using bibliographic references in patents to build linkage schemes. *Scientometrics*, 54(3), 399–420.

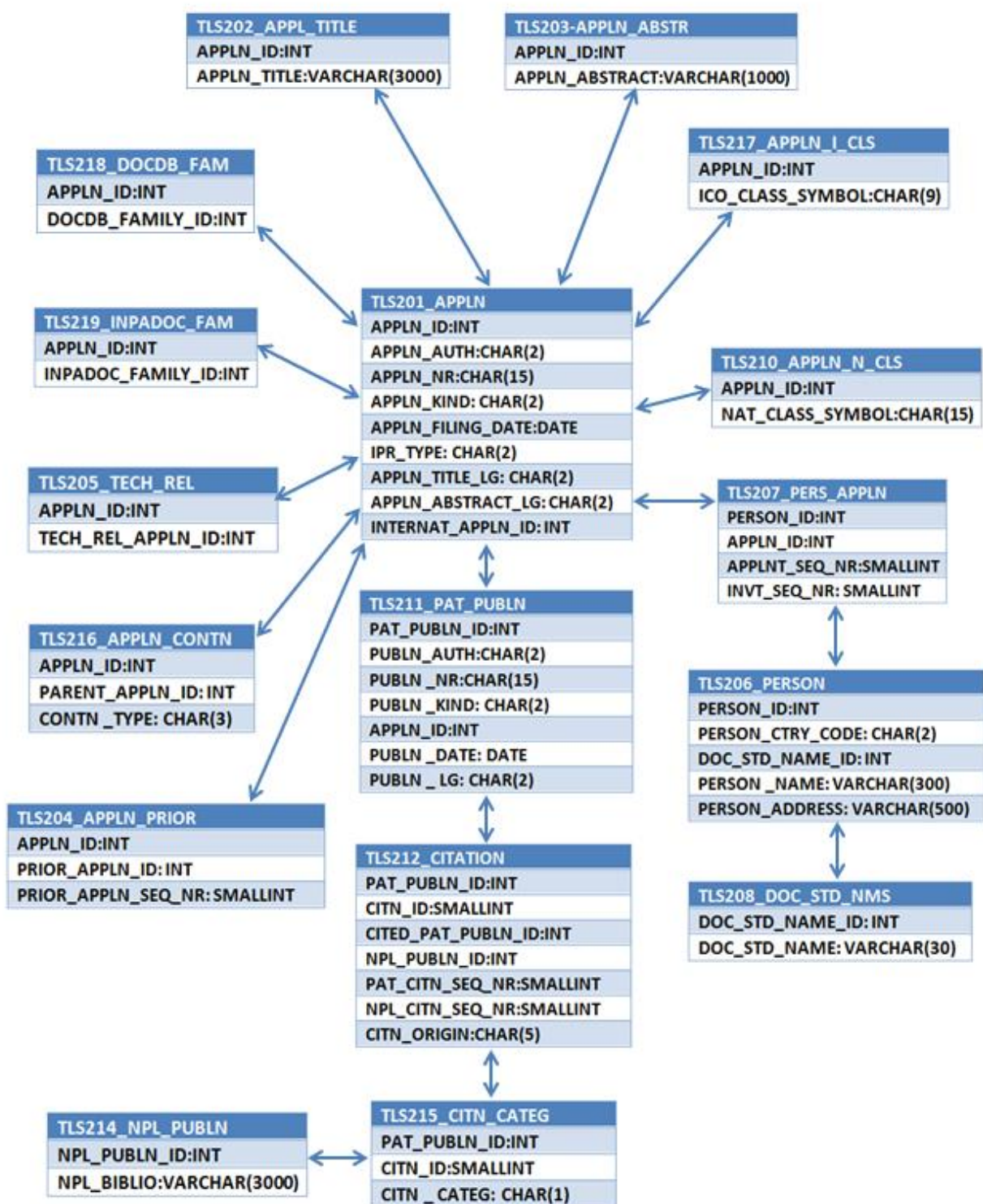
Veugelers R., Callaert J., Song X. & Van Looy B. (2012) The participation of entrepreneurial universities in technology development: Do creation and use coincide? An empirical investigation on the level of national innovation systems. *Economics of Innovation and New Technology*, 21 (5-6), p. 445–472.

---

## **APPENDICES & ANNEXES**



## Appendix 1: Patstat relational diagram



## Appendix 2: Fraunhofer Technology classification (35 fields) (Schmoch, 2008)

The content of each field and the reasons for their specific definition are explained in the following:

1. Electrical machinery, apparatus, energy: the field primarily covers the non-electronic part of electrical engineering, for instance, the generation, conversion and distribution of electric power, electric machines but also basic electric elements such as resistors, magnets, capacitors, lamps or cables. This field is often associated with "traditional" electrical engineering, but the high patent activity shows that technological innovation is still very important.
2. Audio-visual technology: audio-visual technology is largely equivalent to consumer electronics. The relevant IPC codes primarily refer to technologies and only sometimes products are directly addressed (H04R Loudspeakers, H04S Stereophonic systems).
3. Telecommunications: telecommunications is a very broad field covering a variety of techniques and products. The IPC codes are often quite technology-oriented, so that it is difficult to separate relevant product/applications areas such as mobile communication in a clear-cut field. With almost 6 percent of all applications in 2005, telecommunications is one of the largest fields of the suggested classification.
4. Digital communication: in the ISI-OST-INPI classification, this field was part of telecommunications. At present, it is a self-contained technology at the border between telecommunications and computer technology. A core application of this technology is the Internet.
5. Basic communication processes: in the ISI-OST-INPI classification, this field was part of telecommunications. It covers very basic technologies such oscillation, modulation, resonant circuits, impulse technique, coding/decoding. These techniques are used in telecommunications, computer technology, measurement, control. However, the explicit link to these fields by multiple classification is moderate, in the case of telecommunications 2.4 percent. So the definition as a separate field is justified. However, with 0.9 percent of all applications in 2005, it is the smallest fields of the present version of the classification.
6. Computer technology: this field is the largest of the proposed classification with 6.4 percent of all applications in 2005. Its size is already reduced by extracting field 7. The core area of C06F (Electrical digital processing) is defined in a very technical way (Arrangement for programme control, methods and arrangements for data conversion ...), so that a further break-down is difficult. It may be possible to separate specific application fields such as image data processing, recognition of data or speech analysis, but then these special fields may become too small.
7. IT methods for management: a major improvement of IPC-8 is the introduction of the subclass G06Q "Data processing methods, specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes". This field represents software for these special purposes. In most countries, business methods are not patentable, but if they are admitted, they are registered in this subclass. In any case, the size of this field is relevant with 1.2 percent of all applications in 2005. A combination of the fields 3 to 7 represents information technology in general. As the

overlap is limited, this can be done by simple addition. The correct way is to combine the fields without double counting (unit)

8. Semiconductors: the field comprises semiconductors including methods for their production. Integrated circuits or photovoltaic elements belong to this field. The field includes micro-structural technology (B81), as the number of applications in this sub-field is too small for a separate field.
9. Optics: this field covers all parts of traditional optical elements and apparatus, but also laser beam sources. In recent years new optical technologies such as optical switching have become more relevant.
10. Measurement: this field covers a broad variety of different techniques and applications. It would be possible to differentiate special sub-fields such as measuring of mechanical properties (length, oscillation, speed ...), but these sub-fields are generally too small.
11. Analysis of biological materials: this is the largest sub-field of "measurement" and was defined as a separate field. It primarily refers to the analysis of blood for medical purposes. In many cases, biotechnological methods are addressed.
12. Control: In the ISI-OST-INPI classification, this field was part of measuring & control. In recent years the part of control has become quantitatively more important, so that an independent field is justified. The field covers elements for controlling and regulating electrical and non-electrical systems and referring test arrangements, traffic control or signalling systems etc.
13. Medical technology: Medical technology is generally associated with high technology. However, a large part of the class A61 refers to less sophisticated products and technologies such as operating tables, massage devices, bandages etc. These less complex sub-fields represent a large number of patent applications, and the total field is the second largest of the suggested classification with 6.3 percent of all applications in 2005.
14. Organic fine chemistry: without further limitations, the applications in organic chemistry primarily refer to pharmaceuticals. More than 40 percent of the applications have an additional code in pharmaceuticals. As such a large overlap of fields is less appropriate for a classification system, all documents with co-classification in A61K were excluded. The major exception is the group A61K-008, which refers to cosmetics.
15. Biotechnology: biotechnology is defined as a separate field, although it is linked to a variety of different applications. Like organic chemistry or computer technology, it is a crosscutting or generic technology. However, the overlap with pharmaceuticals is too large, with a share of nearly 30 percent. Therefore, as in organic chemistry, applications with explicit co-classification in A61K are excluded.
16. Pharmaceuticals: this field refers to an area of application, not a technology. However, the key subclass A61K is primarily organized by technologies (e.g., medicinal preparations containing inorganic active ingredients ...). Cosmetics are explicitly excluded from the field; these represent about 10 percent of all applications classified in A61K.

17. Macromolecular chemistry, polymers: this field contains the chemical aspects of polymers. Machines for producing articles from plastics are classified in B29 and not included.
18. Food chemistry: this field represents 1.3 percent of the applications in 2005 and is one of the smallest fields in this classification. However, the growth of this field is remarkable, so that a higher weight can be assumed for the next years. Machines for food production are not included, but classified as part of field 28 (other special machines).
19. Basic materials chemistry: This field primarily covers typical mass chemicals such as herbicides, fertilisers, paints, petroleum, gas, detergents etc.
20. Materials, metallurgy: This field covers all types of metals, ceramics, glass or processes for the manufacture of steel.
21. Surface technology, coating: The coating of metals, generally with advanced methods represents the core of this field (C23). Furthermore it covers electrolytic processes, crystal growth and apparatus for applying liquids to surfaces. This field may be qualified as the high-tech part of field 20.
22. Micro-structure and nano-technology: This field covers micro-structural devices or systems, including at least one essential element or formation characterised by its very small size. It includes nano-structures having specialised features directly related to their size.
23. Chemical engineering: This field covers technologies at the borderline of chemistry and engineering. It refers to apparatus and processes for the industrial production of chemicals. Some of these processes may be classified as physical ones.
24. Environmental technology: This field covers a variety of different technologies and applications, in particular filters, waste disposal, water cleaning (a quite large area), gas-flow silencers and exhaust apparatus, waste combustion or noise absorption walls. However, it is not possible to define measuring of environmental pollution by IPC codes in a clear cut way.
25. Handling: This field comprises elevators, cranes or robots, but also packaging devices. So in terms of research intensity, the field is quite heterogeneous.
26. Machine tools: The field is dominated by patent applications referring to turning, boring, grinding, soldering or cutting with a focus on metals.
27. Engines, pumps, turbines: This field covers non-electrical engines for all types of applications. In quantitative terms, applications for automobiles dominate.
28. Textile and paper machines: The fields 27 and 28 cover machines for specific production purposes. Textile and food machines represent the most relevant part of these machines and are classified separately.
29. Other special machines: see field 26.
30. Thermal processes and apparatus: The field covers applications such as steam generation, combustion, heating, refrigeration, cooling or heat exchange.

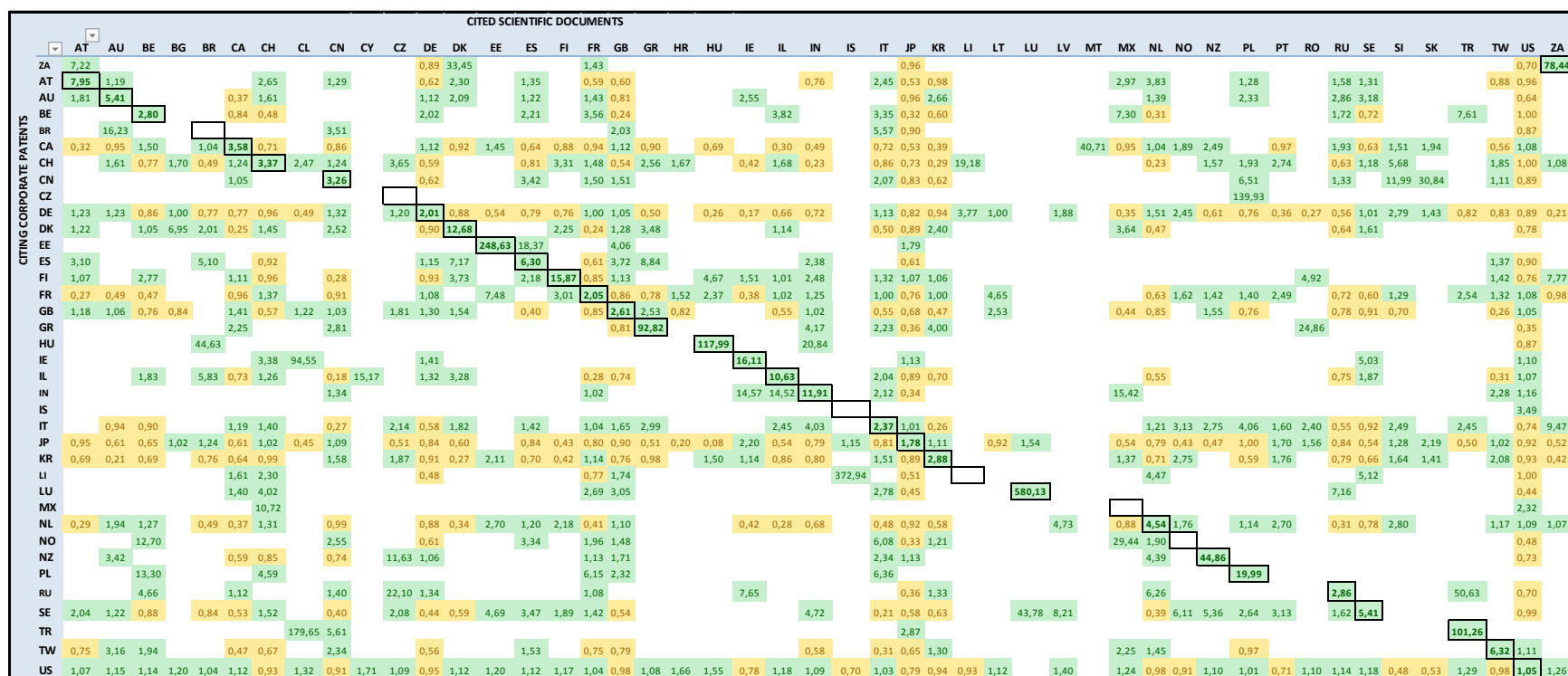
31. Mechanical elements: The field covers fluid-circuit elements, joints, shafts, couplings, valves, pipe-line systems or mechanical control devices. The focus is on engineering elements of machines such as joints or couplings.
32. Transport: the field covers all types of transport technology and applications with dominance of automotive technology. In principle, a separation of rail traffic and air traffic would be feasible, but the associated fields would be too small. In both cases, this is due to a low propensity to patent. The samples are quite small and not representative of the total technological activities in these sub-fields.
33. Furniture, games: this field represents the main parts of consumer goods in terms of the number of patent applications. The other consumer goods are a mix of many different technologies, all of them with low quantitative weight. Therefore a further differentiation is not useful. Even furniture and games combined comprise not more than 2.3 percent of all applications in 2005.
34. Other consumer goods: this field primarily represents less research-intensive sub-fields.
35. Civil engineering: the field covers construction of roads and buildings as well as elements of buildings such as locks, plumbing installations or strongrooms for valuables. A special part refers to mining which may be important for some countries. In general, the importance of mining is so low that the definition of a separate field is not justified.

## Appendix 3: Cross-country science-technology linkage matrix, based on absolute numbers of citations

CITED SCIENTIFIC DOCUMENTS																																			Grand Total			
AT	AU	BE	BG	BR	CA	CH	CN	CY	CZ	DE	DK	ES	FI	FR	GB	GR	HU	IE	IL	IN	IS	IT	JP	KR	LI	NL	NO	NZ	PL	PT	RU	SE	SI	TW		US	ZA	
AT	303	95	63	2	17	156	302	63	1	11	387	51	88	47	173	330	18	26	16	30	22	2	171	291	41		125	18	24	15	13	37	109	9	26	1.012	10	4.104
AU	84	785	94	9	57	271	203	122	1	15	493	92	106	76	322	549	22	23	33	87	60	3	209	510	77		210	47	61	38	13	55	189	2	45	1.600	18	6.581
BE	82	186	548	7	60	342	258	156	2	21	601	97	185	70	547	766	36	39	28	96	61	1	320	581	103		348	43	29	49	19	70	198	2	55	1.808	29	7.843
BG	1	2	7			4	3	3		1	10	3	3		9	10		2		2			7	9	1		8	2					3			18		109
BR	2	5	5		42	8	6	4			18	1	4		21	20	3	2		1	1		14	23	2		5	1	2	1	1		7		1	60	2	262
CA	181	458	358	15	114	2,065	516	317	1	30	1,268	246	378	160	927	1,417	59	42	51	260	155	3	599	1,302	289		575	118	70	68	27	226	439	15	208	4,241	81	17,279
CH	297	504	315	19	98	817	1,565	338	1	62	1,713	288	388	201	1,014	1,736	72	70	101	336	181	8	815	1,540	258	3	594	98	102	103	35	147	555	21	195	4,899	69	19,558
CN	24	45	46	2	12	84	43	182		8	142	23	26	15	105	192	11	3	8	28	29	2	83	212	49		38	17	12	5	1	11	31	2	43	692	6	2,232
CY		2	2			3	1	9			16	2	3	2	8	7	4		4	2		4	9	1		9				2	1		4		2	27		124
CZ	1	2	1		1	3	2	7		12	9		4		4	6	1			1	3		2	15	1		2			2	1	1	3	1	2	35	6	128
DE	819	1,198	1,016	60	213	2,106	1,736	1,094	8	194	8,693	709	1,068	483	3,067	4,566	195	214	227	575	584	24	1,873	4,772	868	11	1,888	229	208	322	156	543	1,307	55	504	14,342	132	56,059
DK	106	173	214	10	34	352	273	116		34	655	809	178	118	453	723	26	28	58	80	99	6	283	603	140		356	61	35	66	40	37	300	6	65	1,757	38	8,332
ES	37	71	54	2	26	107	97	83		9	213	34	461	40	231	314	17	11	10	51	89	2	200	255	65		145	14	20	27	11	17	74	2	35	736	14	3,574
FI	39	80	80	2	22	202	102	83	1	7	307	40	59	456	158	289	44	39	21	43	35	5	128	253	101		120	30	17	12	4	34	158	58	1,294	12	4,335	
FR	172	483	473	29	116	872	588	497	2	54	1,609	240	478	223	2,840	1,735	138	74	86	272	273	6	926	1,834	424	6	606	228	98	129	67	155	432	13	263	5,833	46	22,320
GB	236	557	449	18	125	1,021	791	381		95	1,889	289	511	202	1,347	3,733	111	86	108	271	265	7	924	1,959	391		774	129	130	137	65	207	566	28	241	6,384	103	24,530
GR	1	1	2		3	9	2	2			5	1	2		4	9	18	2		3			11	4	6		2						1		24		114	
HU	5	6	7		3	9	18				26	6	9	6	18	30	1	31	2	6	8		24	31	7		11	3				3	7		120	5	402	
IE	27	18	58	1	8	72	52	34		7	118	19	32	23	67	149	6	9	88	20	28	2	69	98	30		38	17	3	13	3	19	46	17	397	2	1,590	
IL	63	147	138	1	33	311	187	87	3	14	465	74	136	37	308	494	37	23	16	455	95	3	279	573	131		186	28	30	26	5	68	164	5	80	1,606	42	6,350
IN	32	54	66	4	22	129	86	82		23	183	54	81	28	156	305	5	8	12	36	281	7	105	277	81		82	11	15	20	6	15	63	8	35	757	43	3,172
IS	5	10	10	2		15	17	6		4	31	14	12	8	21	43	5	4	2	1	2	44	23	23	3		13	6	1	1		3	20		114	1	464	
IT	52	123	117	7	26	271	205	120		18	465	57	144	58	357	536	34	22	36	96	96	4	952	594	123		217	53	25	41	22	53	138	5	58	1,661	35	6,821
JP	528	1,241	998	62	334	2,746	2,133	1,669	4	176	5,906	704	1,227	567	3,758	5,271	328	169	387	789	677	24	2,000	18,186	2,120	3	2,049	281	176	403	198	684	1,362	58	1,145	23,032	139	81,534
KR	53	154	134	5	23	373	272	312	1	26	822	89	133	125	417	630	37	27	40	122	84	2	261	1,255	1,208		303	73	23	28	16	89	196	11	226	3,055	14	10,639
LI	3	6	14			13	13	14			37	5	6	2	23	23	6		3	5		3	16	48	2	6	10	4				1	10		8	112		393
NL	199	346	361	8	88	652	568	303	4	55	1,526	164	326	235	1,064	1,465	91	51	84	213	176	6	540	1,638	412	1	1,768	102	75	100	73	117	343	10	184	5,240	55	18,643
NO	10	43	23	4	12	97	32	34		7	129	31	38	23	80	188	5	3	6	18	18	12	65	117	18		65	169	2	14	6	10	56	7	19	413	6	1,780
NZ	15	71	26		10	80	34	13		9	89	38	31	22	72	111	2	1	8	16	12		51	107	14		51	8	192	7	4	4	43	2	7	297	1	1,448
PL	2	6	1			8	5	8		2	22	3	3	2	11	11	3	1		1			7	26	10		3			35	1		3	1	69	1	245	
PT		6	4		1	10	7	5			18	3	9	9	21	20			1	3	6	1	15	17	3		5	2	3	1	23	2	7	1	1	53	1	258
RU	1	2	3		1	6	4	8		1	19		4	2	8	18			1	3	4		6	16	8		5	1	1			30	2		49		203	
SE	125	239	233	6	40	424	341	189		33	787	159	261	149	589	1,032	42	35	50	111	122	3	409	837	168	1	291	61	75	59	18	75	1,150	7	119	2,941	23	11,204
SI	6	1	2		1	4	4				9		2	1	14	13		1	2		11	2	9	21	6		1	1	1	1	3		10	25	1	65	3	219
TW	18	35	80	1	8	74	30	82		1	231	12	50	11	88	187	11	4	5	52	96		155	358	150		61	10	1	7	3	9	30		231	902	1	2,994
US	4,629	10,675	8,550	464	2,800	21,396	14,656	8,101	78	1,171	33,155	5,852	7,963	4,855	23,049	34,325	2,239	1,663	1,948	8,391	4,485	170	15,848	36,210	9,151	100	13,995	2,823	1,947	2,547	1,159	4,168	12,232	556	6,410	120,894	928	429,583
ZA	2	5	2	1	3	11	1	3	1	1	26	6	8		10	29		2	2		5		13	25	5		6	1				2	5		54	21	249	
Grand Total	8,160	17,835	14,554	741	4,353	35,123	25,153	14,527	107	2,101	62,092	10,215	14,417	8,256	41,361	61,282	3,627	2,715	3,440	12,474	8,069	352	27,416	74,629	16,467	131	24,964	4,689	3,378	4,280	1,995	6,893	20,262	852	10,285	206,593	1,887	755,675

Overall counts of citation links between countries of corporate patents (appln. years 2000-2009 - EPO, WIPO & USPTO) and countries of cited scientific publications (WoS, 1991-2009)

## ANNEX 1a: Cross-country citation matrices for corporate patents citing scientific documents by FhG-35 technology class<sup>12</sup>



Relative intensities of citation linkages formed by corporate patents (all patent offices) citing scientific publications produced in period 1991-2009 (FhG-1: electrical machinery, apparatus, energy).

<sup>12</sup> Note that rows from citing countries or columns for cited countries were discarded whenever links from or towards them in the technology-field were non-existing.

























[illegible]





















		CITED SCIENTIFIC DOCUMENTS																																					
	AT	AU	BE	BG	BR	CA	CH	CN	CZ	DE	DK	EE	ES	FI	FR	GB	GR	HU	IE	IL	IN	IT	JP	KR	LI	MX	NL	NO	NZ	PL	PT	RU	SE	SI	SK	TR	TW	US	ZA
CITING CORPORATE PATENTS	AT																										9,75												1,71
	AU		5,61			3,22		1,10		0,83						0,81							0,73																2,00
	BE							13,20																															1,71
	BG																					24,81																	
	CA						3,14	0,57	2,14		1,35			12,35										2,34				2,11							5,72				0,83
	CH							4,70			11,37	2,21				0,77	0,72						0,32												3,14				1,14
	CN							5,29			38,38	1,25					4,85																						0,86
	CY																						4,36																1,71
	DE		0,29	0,21	1,17		0,51	0,72	1,32	1,56	0,95	1,55	2,03		0,71	2,03	0,45	0,90	2,54				0,39	0,46	0,95	1,53		1,88	6,10	1,27	1,45	6,10	0,34	1,18			0,19	0,83	
	DK			2,40				3,78	3,77			23,39						2,77			4,74	4,50					14,62								58,48				0,61
	ES							1,92				0,91			20,77	14,88	1,90	1,77						0,79	2,63			2,54									5,51	0,31	
	FI										1,30					14,24						5,48			0,38				2,54										1,94
	FR		1,27		15,45		1,46	0,80	1,99		0,94			2,16		2,76	0,37				1,25		1,40	0,33	1,09									0,80			2,29	1,29	
	GB			1,94			2,38	0,98	2,03		0,31	5,04		3,51		0,96	1,49	12,59						0,27	0,89			0,90						0,83	2,61			1,21	
	IE																																	21,17				1,71	
	IL							3,02	2,52		0,47			5,44			1,00					3,16			2,08			0,93										0,98	
	IN																4,18	3,88		39,30							81,87										0,69		
	IT		1,43	2,68					1,69		0,21						0,44	0,83						1,58	1,49	1,23		0,41	20,90					1,15				1,68	
	JP	0,49		0,77	5,71	0,57	0,27	0,59	1,11	0,54	0,57	1,71		1,00	1,43	1,06	0,91	1,43	0,71	0,69	0,46		0,78	2,19	0,96		2,14	0,34		4,29	1,22	3,43	0,47	0,30		1,22	0,86	1,27	1,08
	KR	0,56	0,40	0,37		0,97	0,46	1,13	2,19		0,94			0,68	0,97	0,87	0,46		9,69	2,33	0,39	1,49	0,44	0,72	4,10		2,42	1,04			1,38	1,28	0,25				0,36	0,93	
	LI							10,59																														1,71	
	LU																																					3,43	
	NL						3,16	0,86	1,08		0,41				3,34		1,58	16,71					3,04	1,25	0,59		1,59									10,02	1,24	1,05	
	NO						6,45				1,66													1,45			3,25											1,14	
	NZ						2,58				1,33					1,39	5,18							0,58			2,60			27,29								0,69	
	RU										3,32																											1,14	
	SE		1,25				0,72	0,39	0,49		1,11	6,06		2,12		1,16	0,36						1,38	0,32	1,07				7,58								1,12	1,27	
	TW		2,59				1,49		5,08		1,15					1,61								0,34														7,00	1,32
	US	1,33	1,29	1,13	0,25	1,28	1,12	1,01	0,78	1,01	1,05	0,59	1,48	0,82	0,79	1,07	1,05	0,49	0,86	1,18	1,30	1,29	1,14	0,88	0,84	1,48	0,74	1,07	0,59	0,37	1,05	0,30	1,25	1,00	0,98	1,26	1,18	1,07	0,97

Relative intensities of citation linkages formed by corporate patents (all patent offices) citing scientific publications produced in period 1991-2009 (FhG-22: micro-structure and nano-technology).

		CITED SCIENTIFIC DOCUMENTS																																												
		AT	AU	BE	BG	BR	CA	CH	CL	CN	CY	CZ	DE	DK	EE	ES	FI	FR	GB	GR	HR	HU	IE	IL	IN	IS	IT	JP	KR	LT	LV	MX	NL	NO	NZ	PL	PT	RO	RU	SE	SI	SK	TR	TW	US	ZA
CITING CORPORATE PATENTS	ZA						0,31			0,48			1,53			0,62		0,51	2,86								1,82						2,00				0,71	0,57					1,00	20,75		
	AT	9,07		2,49			1,22	1,16		0,95			1,66	2,44		0,61	1,83	1,25	1,06	6,52							1,74	0,83		30,70		4,14	1,23			4,09	8,37	2,80	0,56					0,33		
	AU	1,05	9,03	0,72	3,45		0,53	0,45		0,27			1,23	1,42		1,41	1,06	1,02	0,82	1,89			0,86				1,52	0,48	1,64		2,40	0,48		6,68			4,86	0,81	0,65	7,64			0,73	0,96		
	BE	1,50	1,02	6,18		4,88	0,25	1,12		2,74			1,25					1,66	0,80	1,35			1,27				0,72	0,69	0,88			0,51	0,82				1,45	1,16			3,14	1,04	0,93			
	BG									17,04																																	1,98			
	BR			2,64		31,20		0,82		2,01			0,32			1,29		2,12	1,12								0,92	1,75					0,87			3,90					0,70	7,22				
	CA		1,79	0,99		0,59	3,03	0,93		1,32			0,69	0,66	1,70		1,09	0,36	0,70	0,84	1,30			0,67	0,29	0,61		0,78	0,71	0,84		0,82	1,31	0,78	1,14	0,73	1,63		1,53	0,78		1,25	1,03	1,36		
	CH	1,53	0,69	0,84	2,00	1,49	0,77	2,42	2,30	0,80			2,34	1,15	1,44		0,31	0,31	0,84	1,31	0,55		1,02	1,14		0,39		0,95	0,88	0,36			0,62	0,66	0,97	1,55		0,35	1,61		1,59	0,64	1,06	2,30		
	CL		11,79			42,43	4,38		196,45										2,54																											
	CN					4,24		1,68		2,73			1,30			16,58	0,88		1,08	1,78	14,08				1,11		0,60					1,77					1,01					0,95	4,91			
	CY									11,36			1,21						2,00																									1,76		
	CZ								8,52			62,55																															0,99			
	DE	1,37	0,82	0,91	0,96	0,64	0,61	0,81	0,37	0,96	1,99	0,56	1,93	0,69			1,09	0,79	1,02	1,01	1,15	0,83	1,31	0,64	0,60	1,14		1,09	0,89	0,72		0,56	1,07	0,53	0,78	1,39	0,66	1,36	0,64	0,68	1,42	0,26	0,21	0,68	0,93	1,38
	DK	0,66	2,98	0,91	4,32	1,07		0,85		0,69		3,79	0,82	13,31			1,77	2,00	1,37	0,90							1,40		0,32	0,70	1,80		1,50	1,34		2,09			0,25	2,04			0,66	1,24		
	ES			1,87			0,91	0,58		0,71			0,23			12,79	2,75	1,13	1,06	4,89						2,21	3,46		1,30	0,62	2,12			1,23				0,84				0,41				
	FI		2,98				1,13		1,38				0,88			0,44	18,66	1,09	0,77			2,20	2,46		4,47		2,53	0,70	2,57			0,30			2,98			0,51	0,82			0,64				
	FR	0,38	0,34	1,80	1,23	0,91	1,17	0,64	1,47				0,36	0,86	0,38	1,26		2,41	0,79	1,01		0,62		0,61	1,91		0,94	0,86	1,75		0,43	1,27	2,04	1,19	1,33	2,12	0,87	0,72	0,70	1,36		0,79	0,91	0,82	1,06	
	GB	0,64	0,50	0,77	1,57	0,52	0,75	0,82	1,20	0,88			0,31	0,97	0,97	2,03	0,91	1,29	1,24	1,84		2,71	0,80	1,49	1,04	0,68		0,80	0,79	0,81		0,73	0,54	1,39	1,01	0,49	1,08	2,95	1,29	1,28			1,67	1,22	1,01	0,60
	GR																																													
	HU																																													
IE			4,72		5,58		0,74		0,90			1,55						0,95	0,67		31,06					4,47	1,42	7,26									1,32					3,96	1,13			
IL		0,77	1,16			1,14	0,36		0,44	22,96	3,25	0,42	3,42			0,57			0,99			6,40				0,82	1,31						1,15	3,68				2,61	0,52			1,67				
IN	3,27		2,24		5,30	0,82	0,70		0,85			0,81	3,29			1,10	1,65	1,35	1,11						0,78	0,87	1,27				3,72	0,74				1,66		7,53	1,26			1,03	15,95			
IT			1,67	2,00	2,97	0,82	0,52	4,59	0,96							1,84	1,23	1,35	1,19						2,65	7,61						2,78	0,69	1,33	1,94	1,24	1,38			1,32		6,36	1,28	2,97	0,78	
JP	0,39	0,43	0,79	1,10	1,09	0,50	0,78		1,44			0,28	0,72	0,52	0,31	1,11	0,49	1,06	0,77	0,43	0,81	1,60	1,70	0,47	0,73		0,68	2,40	1,33		1,54	0,75	0,58	0,76	1,22	1,09	0,44	0,82	0,54	0,70	1,75	0,81	1,34	0,91	0,72	
KR	0,80	0,54	0,28		0,65	0,81	0,43		1,15			1,54	1,04	0,27		1,08	0,41	0,78	0,74	2,89		1,34		0,65	0,85		0,29	0,98	6,88		3,67	0,82	0,44			0,91	1,85	1,55	0,37			0,56	1,11			
MX																															148,99															
NL	1,26	0,34	0,34			1,22	0,72	0,91		0,85	6,80	2,89	0,81	0,68	3,19	1,18	2,79	1,18	1,10	0,45		1,25	0,94	0,41	1,17		1,45	0,73	0,78		2,39	1,09	0,80	2,55	3,40	2,32	0,48	0,70	1,82		3,15	0,35	0,90	3,78		
NO		2,07			7,44	1,54						0,95				2,31		1,27	0,45						3,72	0,97		0,55	0,35	0,89		2,07	14,93			5,17			0,71				1,11			
NZ																7,31			4,24																								0,66			
PL								109,14				1,21														2,21																	1,32			
PT								3,25				0,52														2,98	0,95	4,84															0,57			
RU																										1,53																	0,61			
SE	2,06	1,24	0,71	1,12	1,11	0,40	0,81		0,81		2,63	0,94	1,61		0,23	1,04	0,71	0,73	1,23			0,64	1,11	0,44		0,58	0,65	0,93			0,39	0,74	2,18	1,04	0,77	1,58	1,19	7,73	2,49		0,72	0,48	1,03			
SI	26,13											1,09						1,80	1,27		21,74					0,99																		0,79		
TR																10,96																											0,99			
TW		1,83						6,26				0,67	1,79		0,89		0,74	1,04						1,13			0,81				6,08												12,58			
US	1,05	1,14	1,00	0,98	0,89	1,19	1,11	1,28	0,89	0,99	1,13	0,92	1,01	1,34	0,95	1,04	0,92	1,00	1,08	1,23	0,94	1,06	1,26	0,96	1,64	1,05	0,84	0,87	1,51	1,64	1,00	1,04	1,10	1,03	0,90	0,88	0,90	1,09	1,00	0,94	1,14	1,19	1,02	1,05	0,78	

Relative intensities of citation linkages formed by corporate patents (all patent offices) citing scientific publications produced in period 1991-2009 (FhG-23: chemical engineering).





		CITED SCIENTIFIC DOCUMENTS																																											
		AT	AU	BE	BG	BR	CA	CH	CN	CY	CZ	DE	DK	ES	FI	FR	GB	GR	HR	HU	IE	IL	IN	IT	JP	KR	LI	LT	LU	LV	MX	NL	NO	NZ	PL	PT	RO	RU	SE	SI	SK	TR	TW	US	ZA
CITING CORPORATE PATENTS	ZA					49,26																			3,84																		0,80		
	AT	9,05					3,73	6,79			2,95									63,34				4,02																				0,68	
	AU		4,93				6,52	1,98							1,77					20,15			3,52																				1,98	1,40	
	BE			12,13			6,95				0,69						1,01	13,14				7,88			1,02							5,44						5,70	3,21						0,21
	CA						4,07	1,17				0,38				1,05	4,51							0,28	1,73												1,58	1,90			4,38	1,17	0,71		
	CH		2,30	2,87				1,90	1,15			1,88			2,80	1,55	0,55	3,58					4,30	2,43	2,05	0,56	0,85				0,74		5,86			0,78	0,93				1,73	0,76			
	CN							7,92							7,09									7,04																			1,70	1,20	
	DE		5,85	1,21	1,05		1,52	1,80	0,20	0,61		2,73	1,79	0,45	0,91	1,19	1,42	1,70	1,52			2,48		1,03	0,87	0,56				1,10			0,80		1,52	0,66	1,38	0,47		0,45	0,73	0,76	13,64		
	DK			6,06					4,87				0,80	18,19	36,38			1,17							1,18																		0,49		
	FI								2,07				2,04					3,04														8,15			41,73			8,55							
	FR			0,83	1,91		4,15	1,16	0,55	5,33			0,76			4,78	1,44							0,70	2,37	0,16						1,72		4,67									0,81		
	GB				2,76		1,67		0,96			0,94					4,61								1,71	0,93	1,41						1,24										0,77		
	IE																																										3,20		
	IL						1,10					2,90				2,27	1,22							2,25	0,31																		1,41		
	IN							1,29				1,29				7,09						14,78	8,37	7,04																			0,80		
	JP	0,69	0,21	0,74	1,60	1,07	0,37	0,28	1,03		3,84	1,49	0,32	1,28	0,42	0,38	1,16	2,67					1,28	0,54	0,31	1,69	1,26			9,61		0,11		1,13			0,81	0,70	0,33		1,63	0,94			
	KR			4,15					2,75			1,63					1,60							1,21	4,91																			1,18	
	NL							13,59	1,80	2,18							2,10						2,31		0,79	1,61																	2,18	1,43	
	NO					197,06																												295,58									10,56		
	NZ						5,50					2,07					3,04								1,54																			0,64	
PT																9,46							18,77											107,48											
SE		2,70				4,19		2,22	2,02		1,54					0,32							1,20	1,80	1,99							4,44					9,84			2,52	1,35	0,20			
SI																																											3,20		
TW								7,92							7,09									1,92																			0,80		
US	0,56	1,08	1,02	0,95	0,63	1,05	1,11	0,78	1,43	0,57	0,82	1,24	0,95	1,18	0,86	0,82	0,71	1,43	1,07	1,04	1,05	1,10	0,95	1,00	1,09	1,43	1,43		1,43	1,43	1,15	1,07	1,25	1,01	1,04	1,27	1,15	0,91	1,33	1,43	1,28	0,98	1,07		





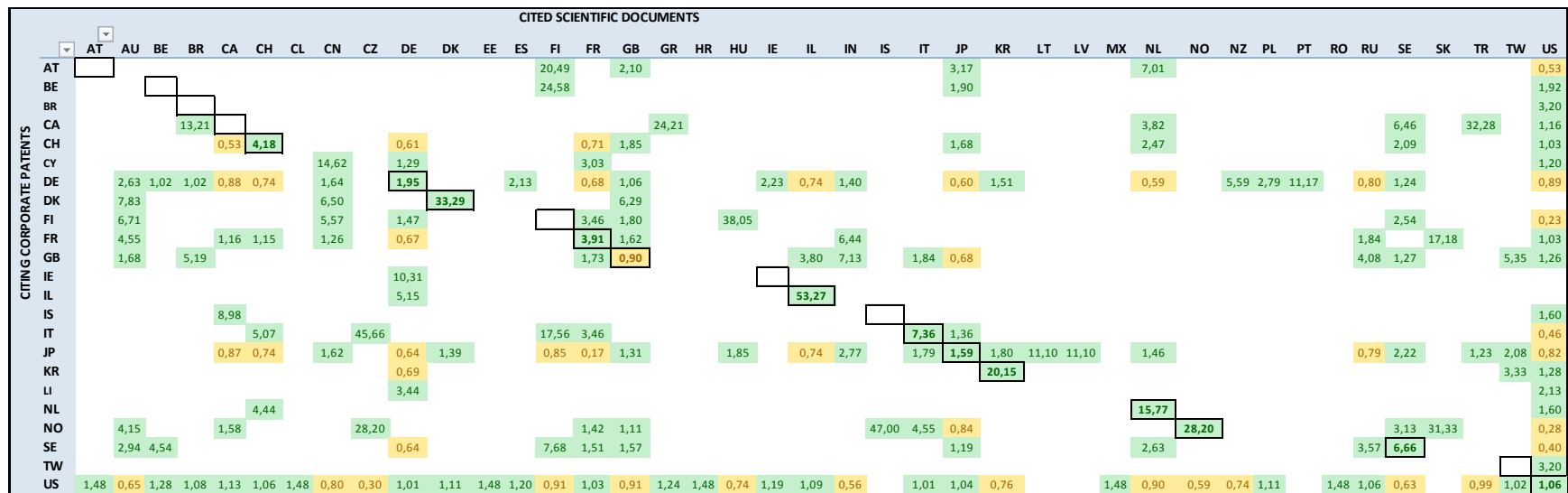
		CITED SCIENTIFIC DOCUMENTS																																								
		AT	AU	BE	BG	BR	CA	CH	CL	CN	CZ	DE	DK	ES	FI	FR	GB	GR	HU	IE	IL	IN	IS	IT	JP	KR	LT	MX	NL	NO	NZ	PL	PT	RO	RU	SE	SI	SK	TR	TW	US	ZA
CITING CORPORATE PATENTS	AT			4,65			1,96										1,13								0,73	4,61						11,34	22,11		5,33	4,09					0,76	
	AU									0,62		0,37						0,47		1,56						0,64				0,25			2,34								2,64	4,89
	BE			2,16			3,63	4,25					1,42	5,87		3,26		0,53						3,09												1,90					1,29	
	CA			1,32			5,53			0,57		1,15	7,14		1,98		0,41	0,64			2,60				0,21				0,68		22,72				1,16			16,66		0,93	8,93	
	CH	2,19		0,81	6,39	2,89	0,34	4,76		0,70		0,88			0,74		0,51	0,98			1,60	1,97		1,15	1,66	1,60			1,25			1,97	7,66		0,71					0,70		
	CN			8,64					7,50								2,72	2,10		28,32					1,36																	
	CY								13,12														4,91			1,27									3,55					5,04	1,32	
	CZ								7,00																																	
	DE		1,00	0,66	0,92		0,66	0,47	0,85	17,52	0,80	1,70	2,50	1,18	0,83	0,87	0,90	1,17	0,60			0,73	2,02		1,58	0,67	0,64	2,92	1,59	1,14		3,19		0,88	2,50	0,63	0,97		0,69	1,11	1,25	
	DK		3,13	1,37	0,58		1,03	0,97	1,89		0,25	10,43	0,76	10,95	5,26	0,87	2,36	1,68							1,00	1,14			1,79	2,11					0,51			4,98	7,30	0,28	3,91	
	ES														9,21				2,45																						1,64	68,43
	FI	5,47		1,01				1,06	1,32				0,66	6,84	0,92	9,88	1,11	1,10	3,19			2,99	1,23		0,36	0,80	1,00		4,35	0,78		1,23			3,55					1,74	3,42	
	FR			0,70	1,76	4,65				2,04	1,03	1,59			1,77	1,66	1,00		1,92		3,49	2,15			1,68	0,56	0,29		1,52	8,59		4,29		1,34	0,52				3,72	1,47	0,57	
	GB							1,78	1,04		0,46		0,82		0,97	0,80	1,17	2,71	6,72	1,74		1,29			0,38	1,26	0,79		1,10	5,82			2,52	14,41	0,61	0,47		3,36			0,75	
	IE								8,81			30,41	2,95							14,52															7,69						0,73	
	IL												0,83					0,92							2,70						13,82										1,23	
	IN												1,77		7,37		1,27	1,96							1,91	2,00															0,88	
	IT									1,09	11,40	0,55					0,79			8,26			9,21		7,20	1,59	2,49								2,89					3,15	0,55	
	JP	0,18	0,64	1,41	1,59	0,36	0,65	1,18		1,48	1,21	0,91		0,37	0,10	0,86	0,99		2,19	1,01	0,13	0,65		0,53	1,90	1,06	2,12	1,45	0,76			0,16	0,95	0,91	0,77	0,77			0,85	0,67	1,00	0,45
	KR	1,87	0,82	1,38	5,44	1,23		0,45		1,79		0,91		0,63		0,87	1,00		2,25	1,48	1,36	2,51		0,49	1,19	2,72		5,94	1,06						2,36		10,89			0,86	0,82	
	NL				4,84	1,10	0,51	0,80		0,27	5,53	0,94	1,66	2,23	3,69	1,54	0,89		2,00		1,21	1,49		1,31	0,68	0,91			1,89					2,10	0,54					1,10		
	NO			3,36			1,41			1,46			9,12	3,07		2,11	0,82							2,40	0,53					1,74	12,28		15,97		2,96					0,73		
	NZ	6,08	2,66	2,24			0,94	1,47				1,47		4,09	3,38		0,54				8,87			1,60	1,06	1,11			2,31		19,35		10,64							0,37		
	SE			2,42	19,16	2,17	1,53	3,17				0,80	3,28			1,14	0,88							0,19	0,60			0,62			5,90		5,54	10,64			7,66			0,53		
	TW		2,99	2,52					1,65	7,11		0,28				1,19			8,26		1,54				0,80																1,23	
	US	1,17	1,38	0,93	0,44	1,41	1,17	0,91		0,88	0,76	1,05	0,56	1,11	0,99	1,08	1,01	1,30	0,61	1,29	1,15	0,87	1,78	1,19	0,85	1,10	0,89	0,89	1,09	1,09	0,97	1,18	0,89	0,76	1,16	1,00	1,48	1,61	0,71	1,40	0,94	0,63

Relative intensities of citation linkages formed by corporate patents (all patent offices) citing scientific publications produced in period 1991-2009 (FhG-28: textile and paper machines).

Relative intensities of citation linkages formed by corporate patents (all patent offices) citing scientific publications produced in period 1991-2009 (FhG-29: other special machines).

		CITED SCIENTIFIC DOCUMENTS																																				
		AT	AU	BE	BG	BR	CA	CH	CN	CY	DE	DK	EE	ES	FI	FR	GB	GR	IE	IL	IN	IT	JP	KR	LI	NL	NO	NZ	PL	PT	RU	SE	SI	SK	TR	TW	US	ZA
CITING CORPORATE PATENTS	ZA															12,61																				1,78		
	AT	29,11					2,97		1,99		1,21											15,59				2,55												2,14
	AU								5,18													3,15																2,14
	BE						12,87										16,82																					
	BR																																		236,50		1,78	
	CA																	10,87					2,63													0,59		
	CH							7,18				1,47			3,67					4,51	1,07		0,73									1,10					1,33	
	CN							9,65	6,48																												0,89	
	DE				0,40			0,39	0,39	2,85		2,13			1,58		0,76	1,47				0,46	1,01	0,79	0,46	18,92	0,66	1,96	6,31	7,57		0,68	1,42			0,68		
	DK				1,66					4,32				8,76								2,82	1,31				8,30	5,44								0,30		
	ES			11,40						12,96																												
	FI		9,96						1,36		0,41	14,23		2,77	11,06		0,86				9,72	1,78	0,83	7,29								2,49				0,38		
	FR			0,38				0,74	1,99		1,52			3,03		3,40	0,31				1,95	0,30			0,64	1,25	12,13				0,91			1,73	1,10			
	GB		7,28					2,97			1,21			2,02		1,94	1,88				1,77	1,30	1,21		1,28			14,55								0,96		
	IL																																		473,00			
	IT										1,21			4,04						3,73										5,20	10,92			36,38	0,82	36,38		
	JP						0,84	0,42	1,69		0,77	5,88		1,71		0,82	0,89				1,84	2,06	1,50		0,36	5,67	6,86	4,11				20,57		0,98	1,08			
	KR																					2,63														15,02	1,19	
	LI																																				3,56	
	LU										3,94																											
	LV																25,23																	23,65				
	NL		1,42							78,83	0,49						1,02										4,15										1,11	88,69
	NO																					2,11																
	NZ																																					
	PL																																					
	PT																																					
RU																																						
SE																																						
SI																																						
SK																																						
TR																																						
TW																																						
US		0,78	1,26	1,24		1,30	1,16	0,93	0,80	0,43	0,94	0,74	1,30	0,68	1,15	0,95	1,03	1,30	1,30	1,13	1,07	0,81	0,99	1,07		1,05	0,71		0,26	1,11	1,20	0,84		1,30	0,32	1,11	1,02	

Relative intensities of citation linkages formed by corporate patents (all patent offices) citing scientific publications produced in period 1991-2009 (FhG-30: thermal processes and apparatus).



Relative intensities of citation linkages formed by corporate patents (all patent offices) citing scientific publications produced in period 1991-2009 (FhG-31: mechanical elements).

		CITED SCIENTIFIC DOCUMENTS																																									
	▼	AT	AU	BE	BG	BR	CA	CH	CL	CN	CY	CZ	DE	DK	ES	FI	FR	GB	GR	HU	IE	IL	IN	IT	JP	KR	LU	NL	NO	NZ	PL	PT	RO	RU	SE	SK	TR	TW	US	ZA			
CITING CORPORATE PATENTS	AT	25,45	7,83																										21,43										0,55	407,20			
	AU					25,45		5,99					1,96		20,36			1,17			16,97																		0,28				
	BE												4,89																														
	CA		3,86	1,19				4,49											1,77					0,85	0,46	1,19													1,17				
	CH		4,54		4,28			8,55						1,75		4,85			0,84					1,99	0,55														0,89				
	CY																																										
	DE			0,40	0,61		1,31	0,76	0,31		1,28	2,09		1,51		0,35	1,49	1,60	0,54	1,49		3,48		1,25	2,57	1,10	0,80		41,55						0,40	1,55			0,20	0,92			
	DK							18,51																					1,49		1,74		2,09						1,38				
	ES															13,57		4,79	2,34					16,29															0,55				
	FI																																										
	FR		1,07	0,66	2,01		6,42	0,62	3,02		1,40		0,74		1,14	4,89	1,81	1,08	1,22	11,41		1,07	0,68	1,87	1,16	1,32		1,05		2,85				0,66	0,63			0,34	0,77				
	GB		7,95					1,16	1,87					3,18			2,25	1,83					2,55			1,22								7,34					1,21				
	IL				13,31													2,60																					1,84				
	IN																							40,72																1,38			
	IT		4,39	2,70	4,13		1,28			1,43			0,68		2,34			0,40						2,81	0,96	1,58	5,40		4,30						2,60				0,57				
	JP		0,68	0,31	1,61	1,82	0,68	1,19	0,64	5,46	0,78		3,28	0,92	1,09	1,46		0,77	0,69	1,56	1,82	0,45	0,34	1,09	0,82	1,44	1,99		1,23		1,56	3,28		0,84	1,62	5,46	4,09	3,42	0,81				
	KR												1,22				5,99					15,91													9,79				4,99	0,69			
	LI							59,88																																			
	LU		31,81	4,89									12,73									31,81																			0,69		
	NL			2,06	6,30					2,19			1,03				2,52	1,23					4,29	2,94					2,19												0,87		
	NO			19,58																																						0,69	
	PL																																										
	RU																																										
	SE						4,63											1,46									4,89											28,28				0,69	
	TW										41,55																																
	US		0,56	1,28	0,53	1,20	0,45	0,95	0,85		0,88	1,44	0,72	1,05	1,26	0,54	1,03	0,87	1,15	0,90		0,90	1,24	0,72	0,71	0,93	0,62	1,80	0,81	1,70	1,20	1,03	0,36	1,80	1,11	0,67		0,45	0,56	1,13			

Relative intensities of citation linkages formed by corporate patents (all patent offices) citing scientific publications produced in period 1991-2009 (FhG-32: transport).

		CITED SCIENTIFIC DOCUMENTS																																				
CITING CORPORATE PATENTS	<div><div></div><div>▼</div></div>	AT	AU	BE	BR	CA	CH	CL	CN	DE	DK	ES	FI	FR	GB	GR	HU	IE	IL	IN	IT	JP	KR	LI	LT	MX	NL	NO	NZ	PL	PT	RU	SE	SI	TR	TW	US	ZA
	AT					10,41																															1,36	
	AU		3,34		17,37	1,81				1,14		3,34	6,20		1,09	43,43					1,00						1,61					21,72	2,55				0,47	
	BE																																				2,71	
	CA					5,48			3,09	1,38	15,02				1,31						3,76																0,86	
	CH																																				2,71	
	CN																				5,74															1,36		
	CY									4,38					4,16						9,25																	
	DE					0,87	3,67			0,55					2,60					8,33	2,31	0,48	2,97												1,39	0,68	13,88	
	DK	#####													1,56			49,95		6,94															0,68			
	FI									4,38			47,57													12,33												
	FR						6,78			1,01		5,91								4,27	1,77						6,40									0,63		
	GB					0,52	0,73			0,99			3,57	1,19	0,94					2,08	1,72	3,57				3,70				5,00	6,24				0,83	0,81		
	IL		12,81				9,79																													0,90		
	IN																																			2,71		
	IT						7,35															2,87													0,68	83,25		
	JP			0,90		0,92	0,93			1,00	0,90			0,90	0,55				1,81	0,35	1,31	1,81			6,32	6,32		3,16		1,05	2,53		2,23		0,21	1,36		
	KR						9,79																													1,81		
	LU																				5,74															1,36		
	NL						1,63		3,26			4,27		2,64	1,39												6,17									1,36		
	PT																																				2,71	
	SE						9,79								4,16																					0,90		
	TW																																				2,71	
	US		1,37	1,34	1,30	0,91	0,98	0,71	1,51	1,33	1,05	0,86	1,16	0,86	1,15	1,06		1,51	0,91	1,08	1,21	1,01	0,96	0,65	1,51		1,01	0,76	1,51	1,13	0,60	0,38	0,89	1,51	1,51	1,36	0,94	0,50

Relative intensities of citation linkages formed by corporate patents (all patent offices) citing scientific publications produced in period 1991-2009 (FhG-33: furniture, games).

		CITED SCIENTIFIC DOCUMENTS																																												
		AT	AU	BE	BG	BR	CA	CH	CL	CN	CZ	DE	DK	EE	ES	FI	FR	GB	GR	HR	HU	IE	IL	IN	IT	JP	KR	LI	LT	MX	NL	NO	NZ	PL	PT	RU	SE	SI	TR	TW	US	ZA				
CITING CORPORATE PATENTS	ZA																																													
	AT											3,22						1,88								1,42																		2,92		
	AU														2,05		0,48	0,63								0,24																		1,10		
	BE										9,66		6,44																															2,56		
	CA							2,18	2,23									1,91	2,50																									0,73		
	CH							1,54	1,58																																			0,73		
	CN																																											0,69		
	DE																																											0,97		
	DK																																											0,42		
	ES																																											0,65	0,75	4,41
	FI																																											0,42		
	FR																																											0,65	0,75	4,41
	GB																																											0,65	0,75	4,41
	GR																																											0,65	0,75	4,41
	IE																																											0,65	0,75	4,41
	IL																																											0,65	0,75	4,41
	IN																																											0,65	0,75	4,41
	IT																																											0,65	0,75	4,41
	JP																																											0,65	0,75	4,41
	KR																																											0,65	0,75	4,41
	LI																																											0,65	0,75	4,41
	LT																																											0,65	0,75	4,41
	MX																																											0,65	0,75	4,41
	NL																																											0,65	0,75	4,41
	NO																																											0,65	0,75	4,41
NZ																																											0,65	0,75	4,41	
PL																																											0,65	0,75	4,41	
PT																																											0,65	0,75	4,41	
RU																	</																													

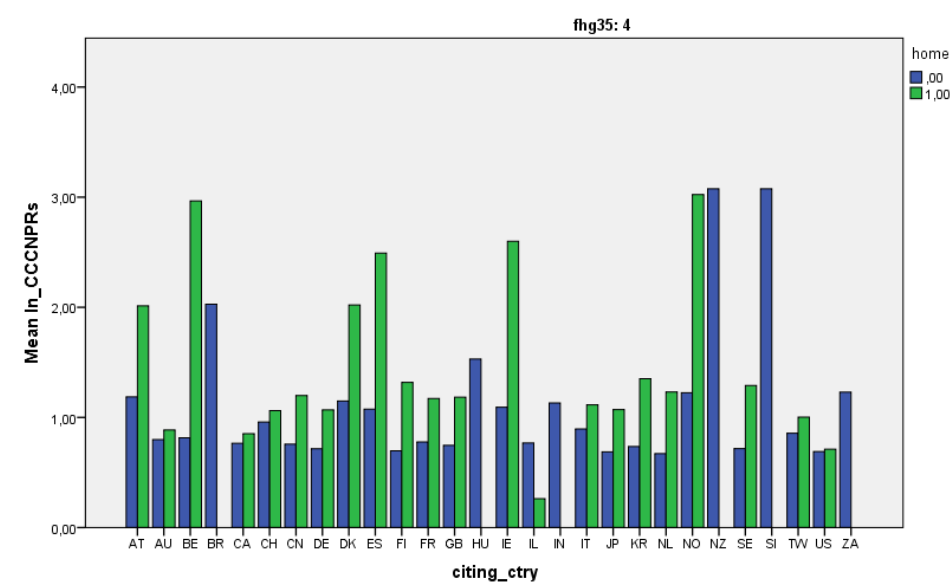
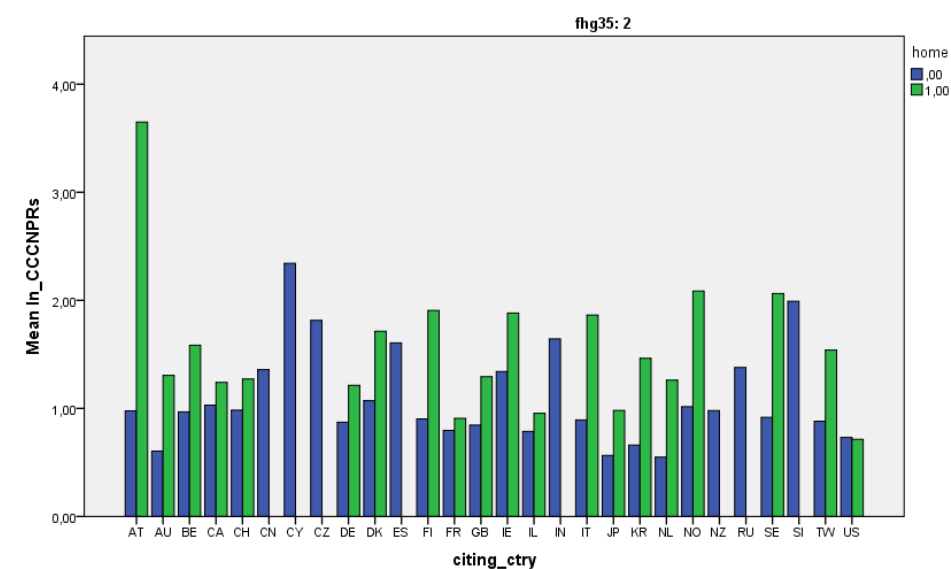
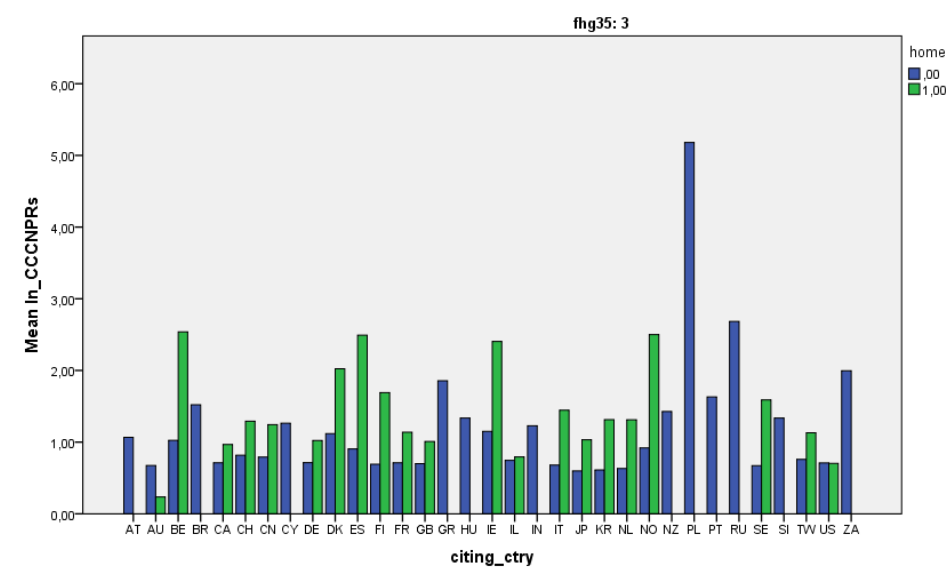
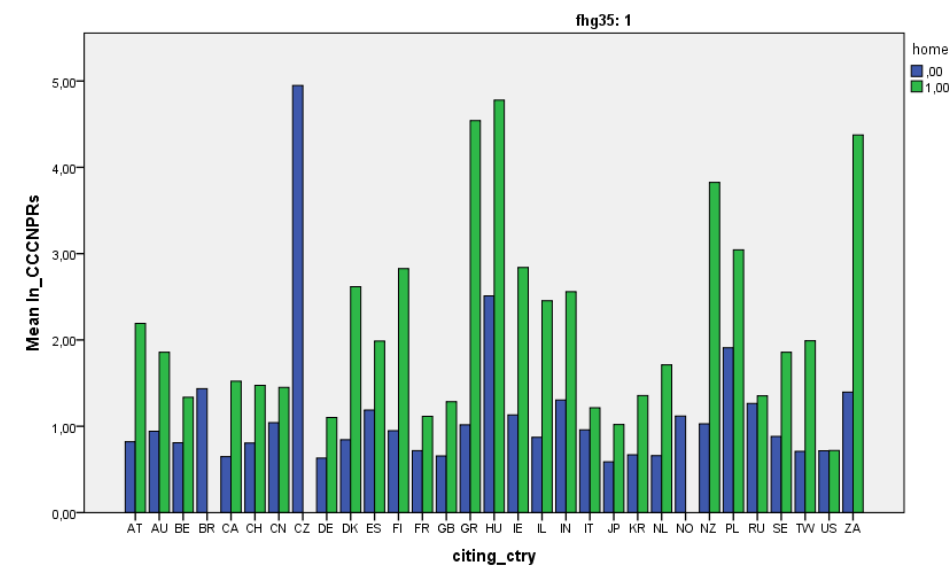
Relative intensities of citation linkages formed by corporate patents (all patent offices) citing scientific publications produced in period 1991-2009 (FhG-34: other consumer goods).

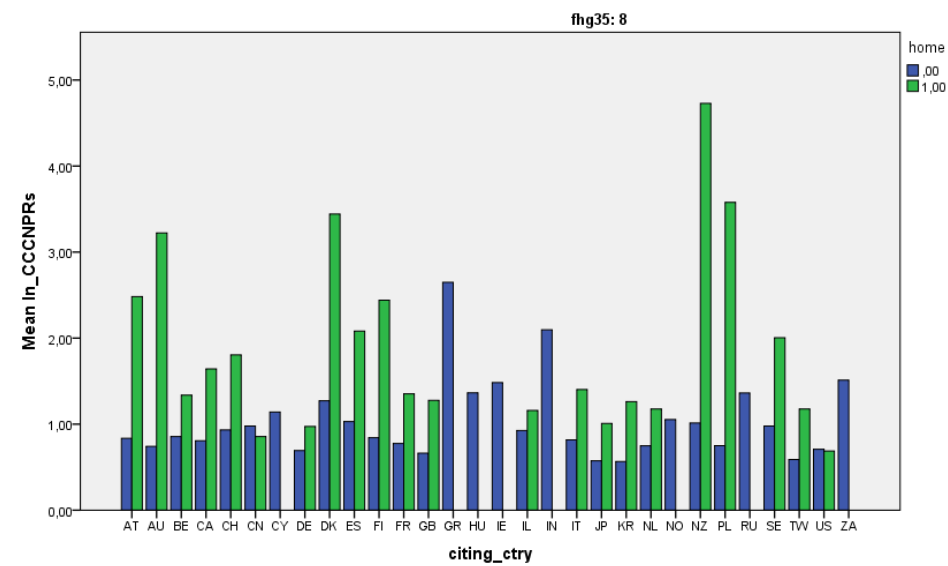
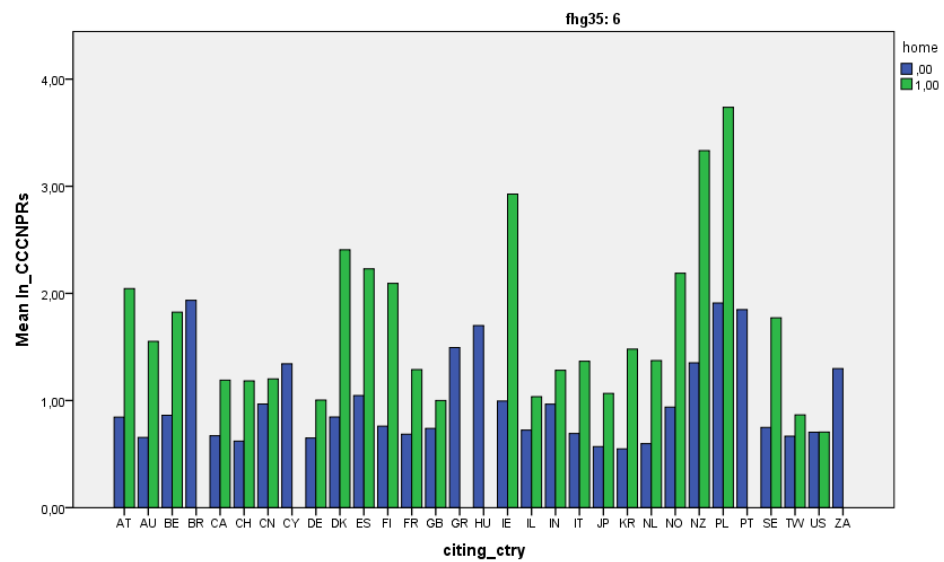
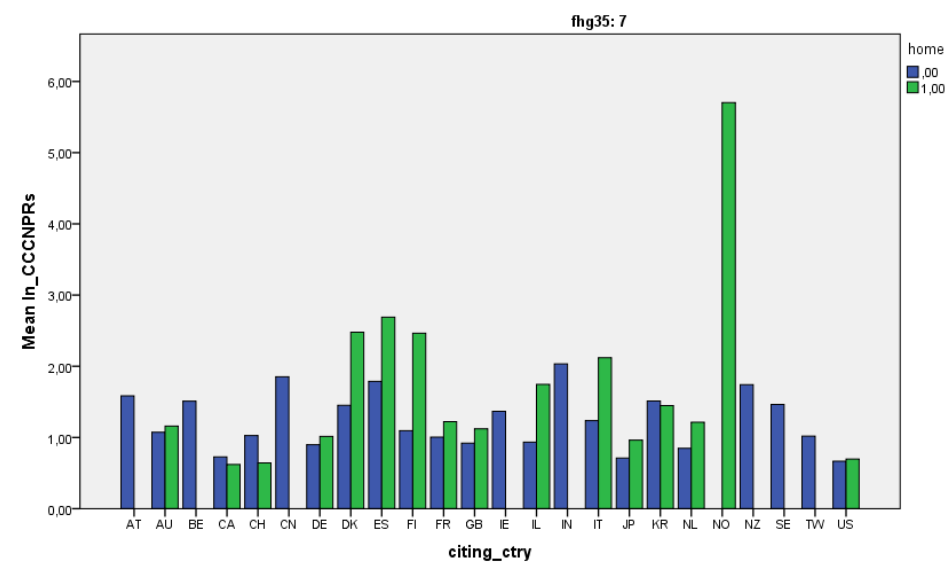
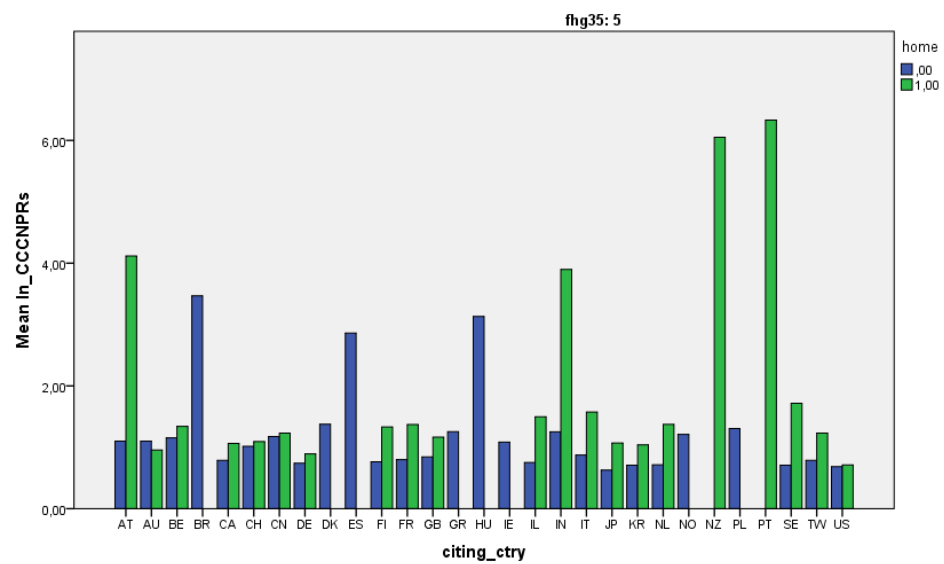


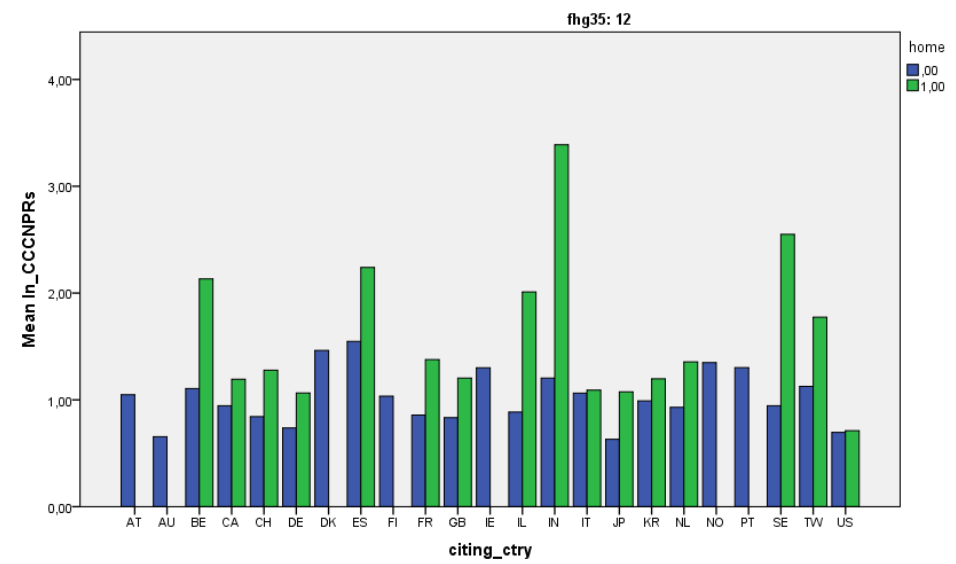
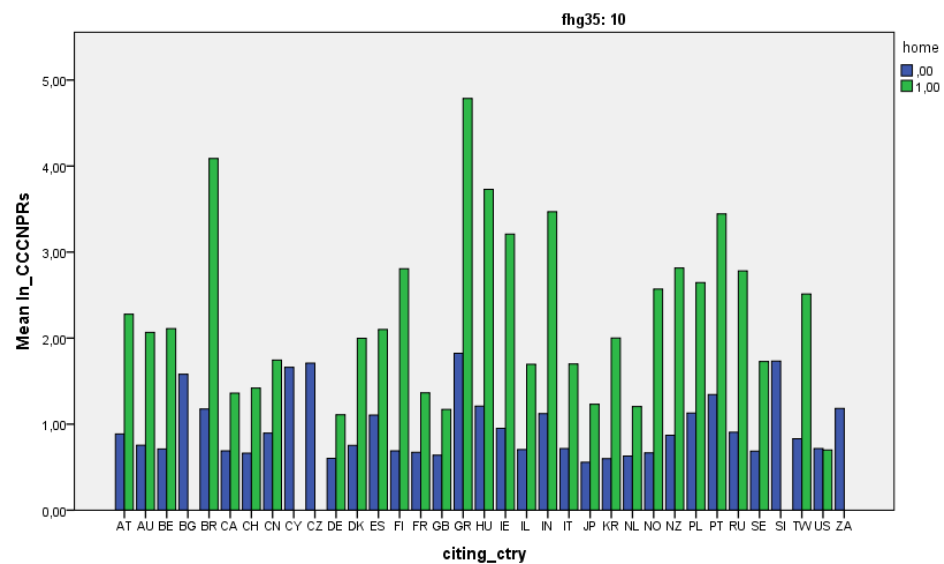
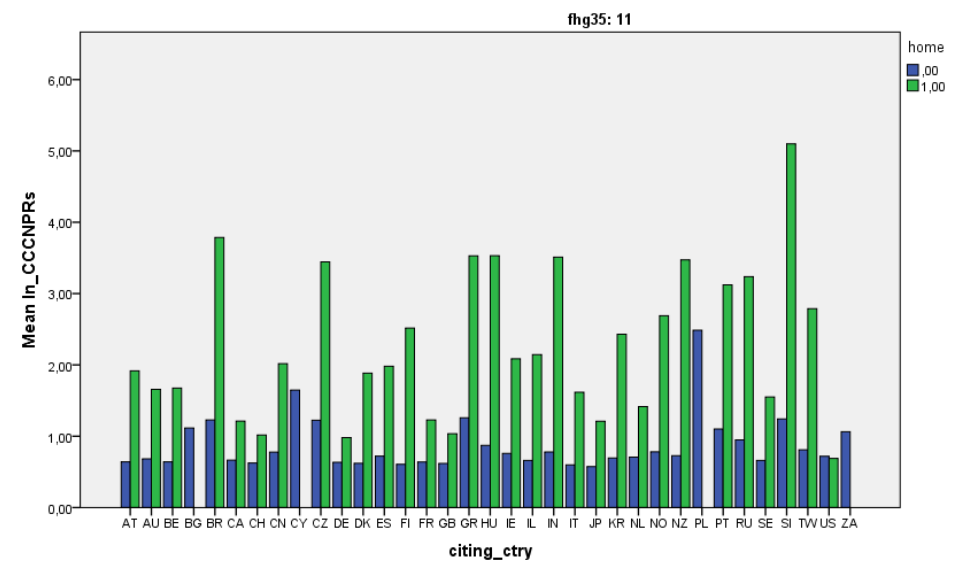
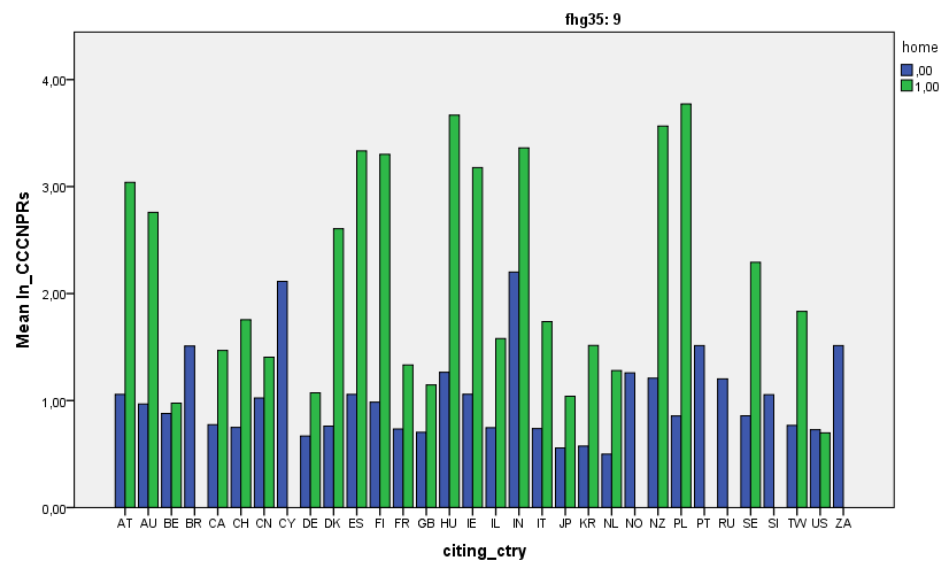
		CITED SCIENTIFIC DOCUMENTS																																							
		AT	AU	BE	BG	BR	CA	CH	CN	CZ	DE	DK	EE	ES	FI	FR	GB	GR	HU	IE	IL	IN	IT	JP	KR	MX	NL	NO	NZ	PL	PT	RO	RU	SE	SI	SK	TR	TW	US	ZA	
CITING CORPORATE PATENTS	ZA																							18,17															0,99		
	AT										7,93																													1,98	
	AU																						3,85		4,54															1,98	
	BE																5,85	2,29						5,45																1,19	
	CA																																							16,50	
	CH																																							0,37	
	DE																																							0,86	
	DK																																							0,99	
	FI																																							2,97	
	FR																																							0,86	
	GB																																							0,68	
	IE																																							0,59	
	IN																																								
	IT																																								
	JP																																								
	KR																																								
	NL																																								
	NO																																								
	SE																																								
	US																																								

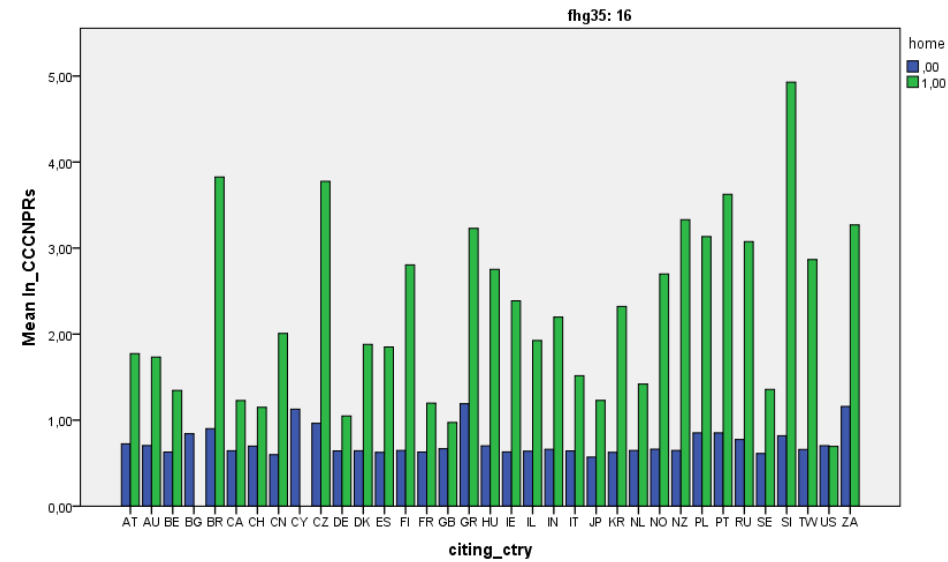
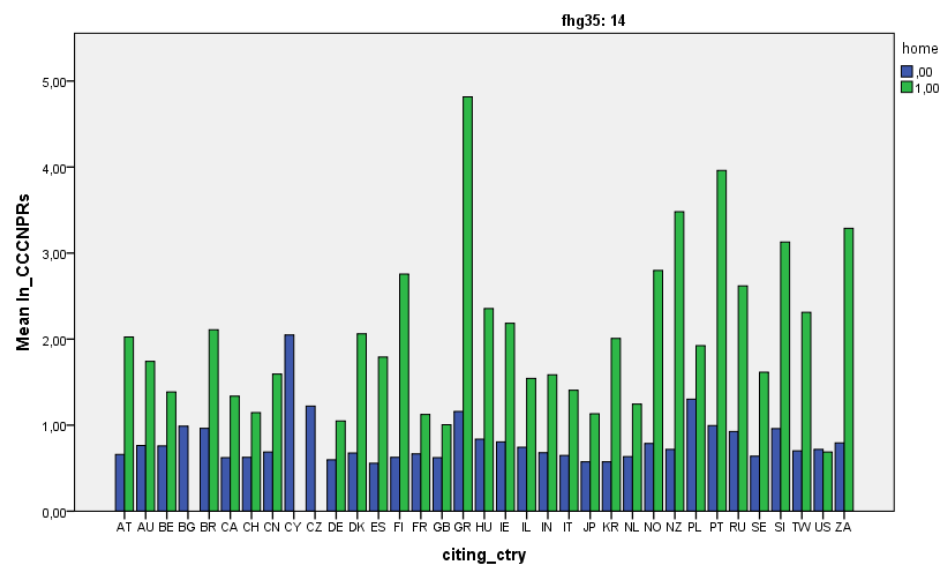
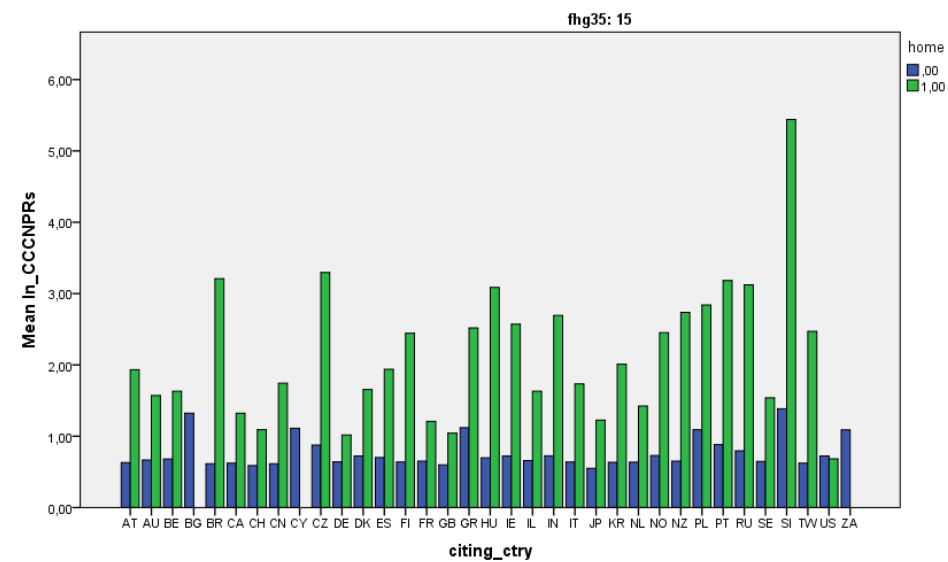
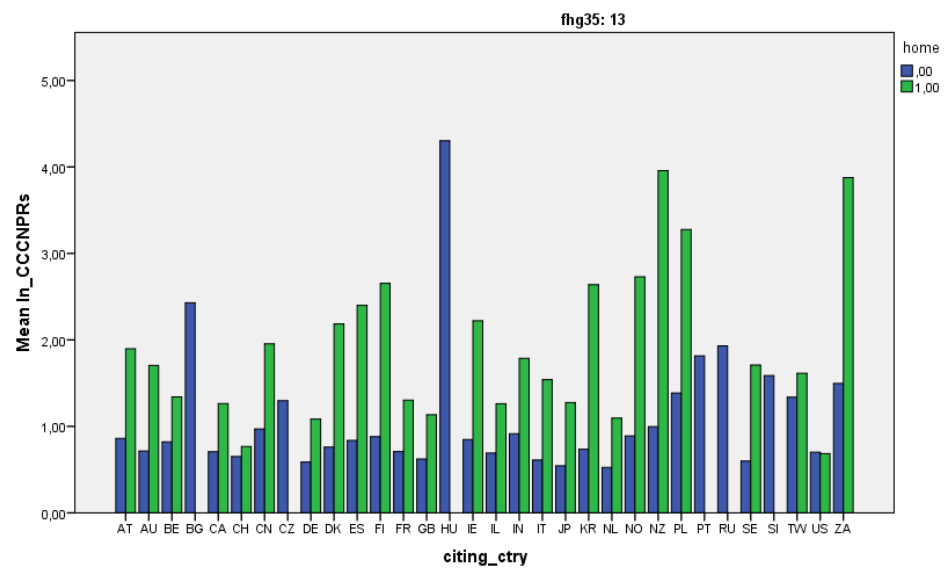
Relative intensities of citation linkages formed by corporate patents (all patent offices) citing scientific publications produced in period 1991-2009 (FhG-35: civil engineering).

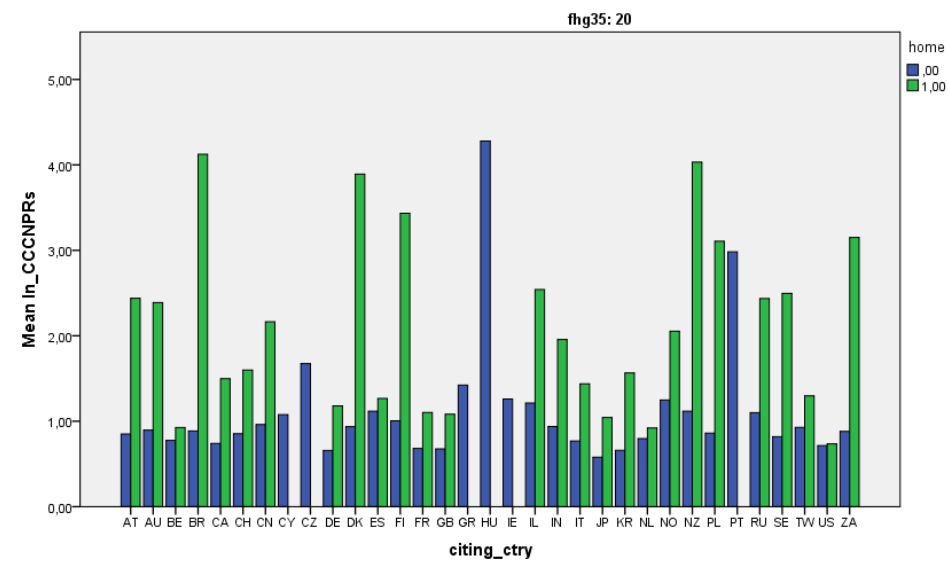
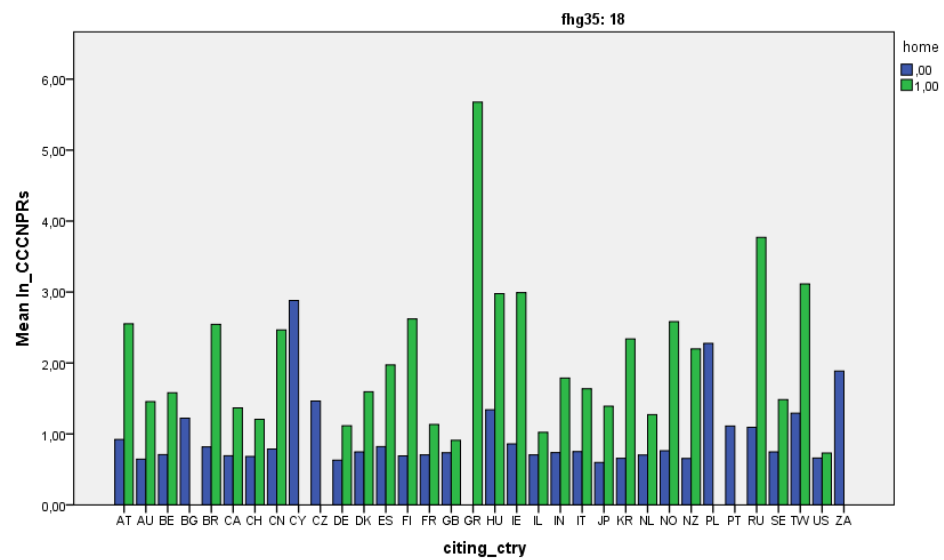
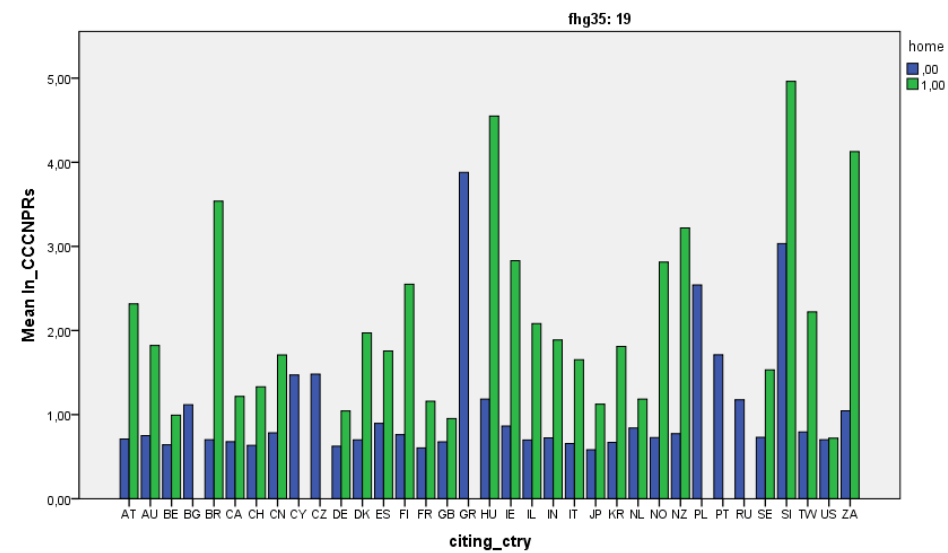
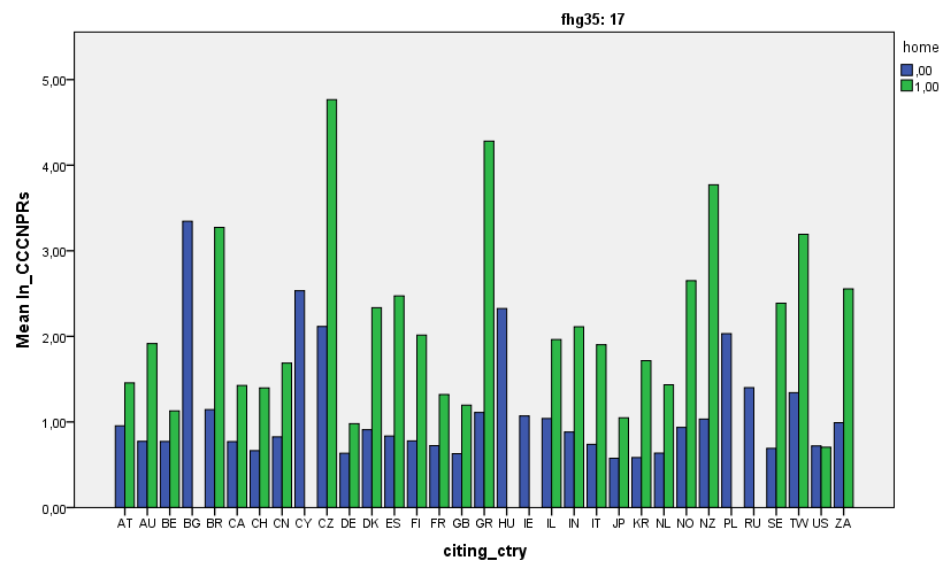
ANNEX 1b: Comparison of within-country (home) citation intensity (*green*) to the average citation intensity with foreign countries (*blue*), by FhG-35 technology class

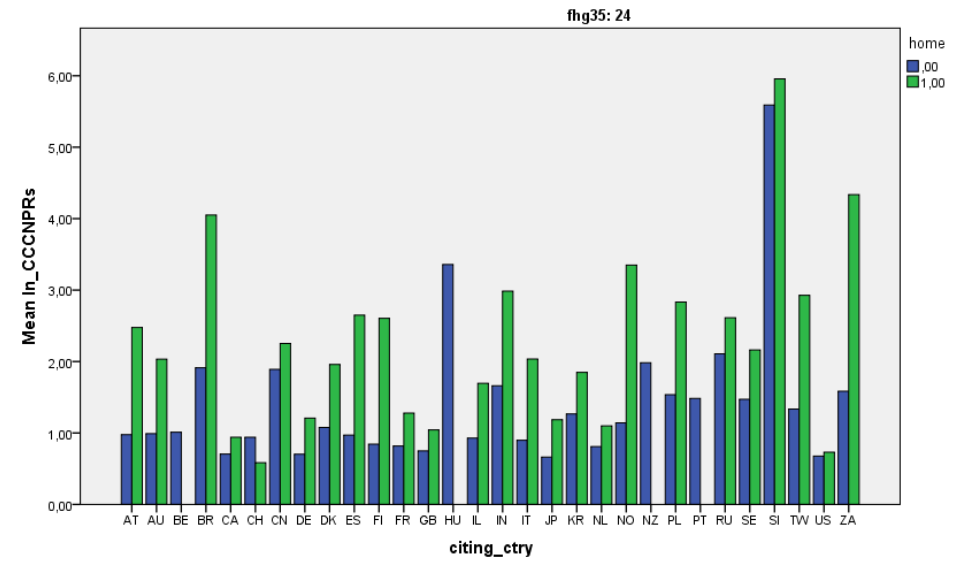
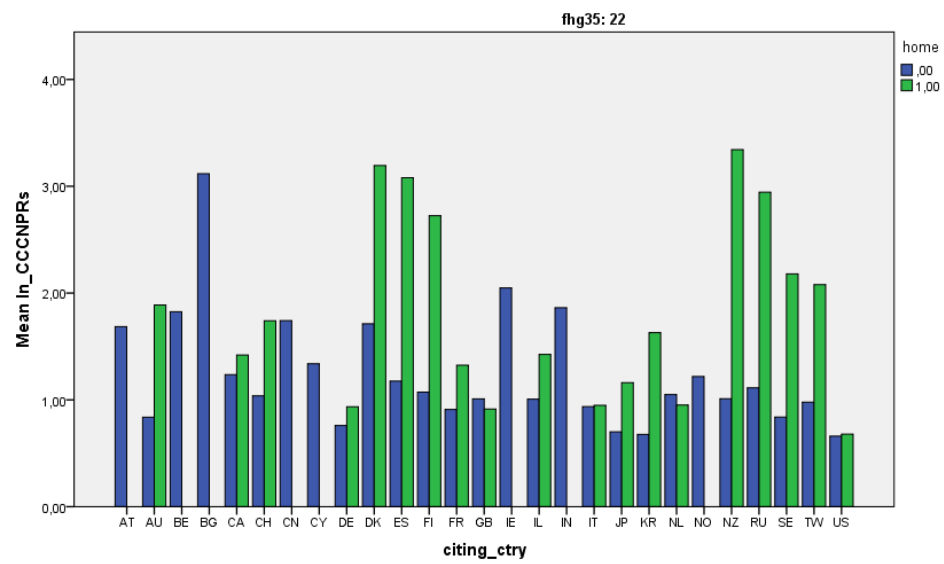
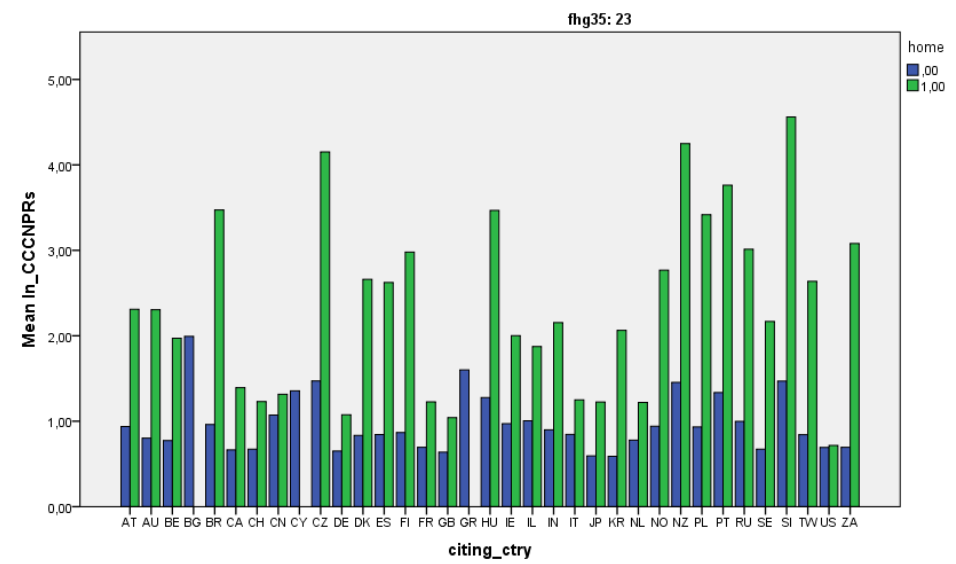
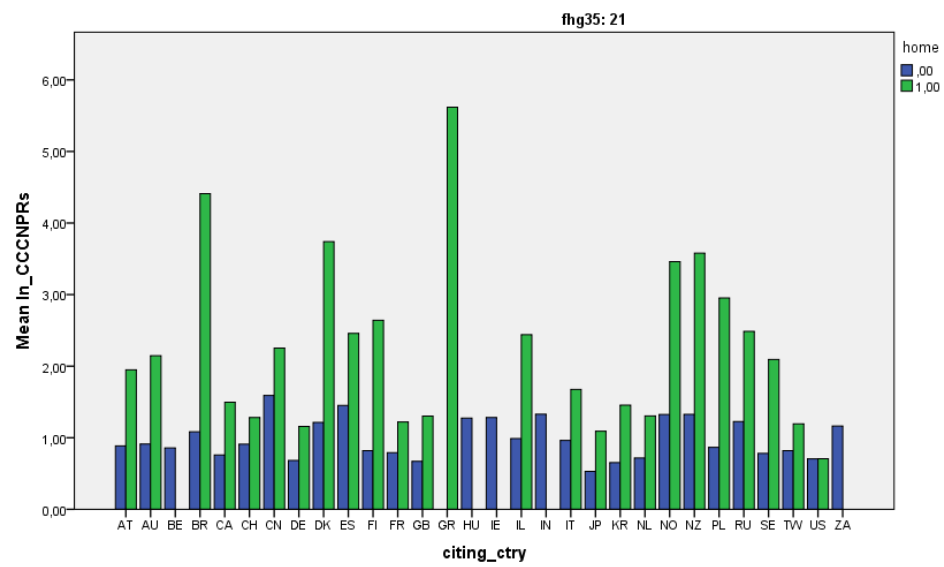


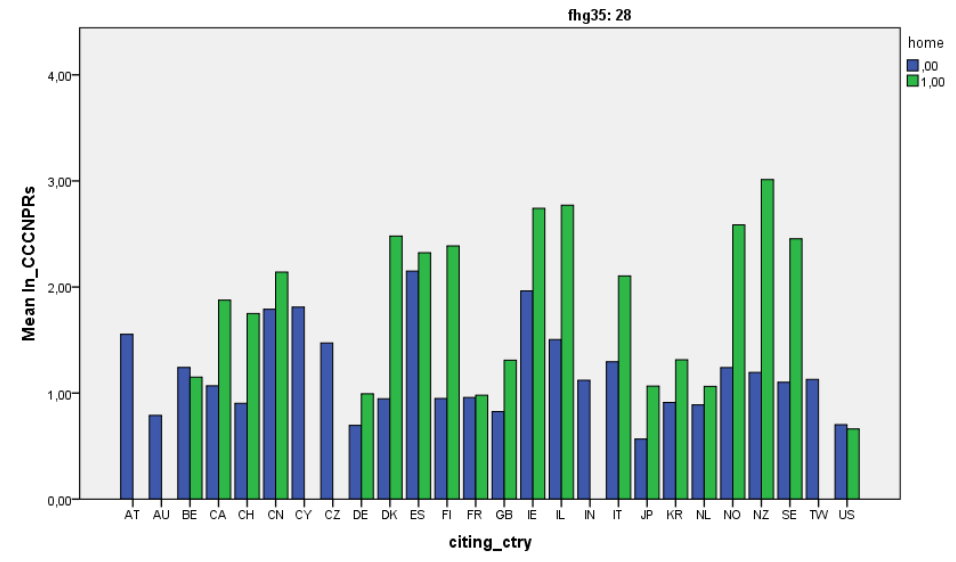
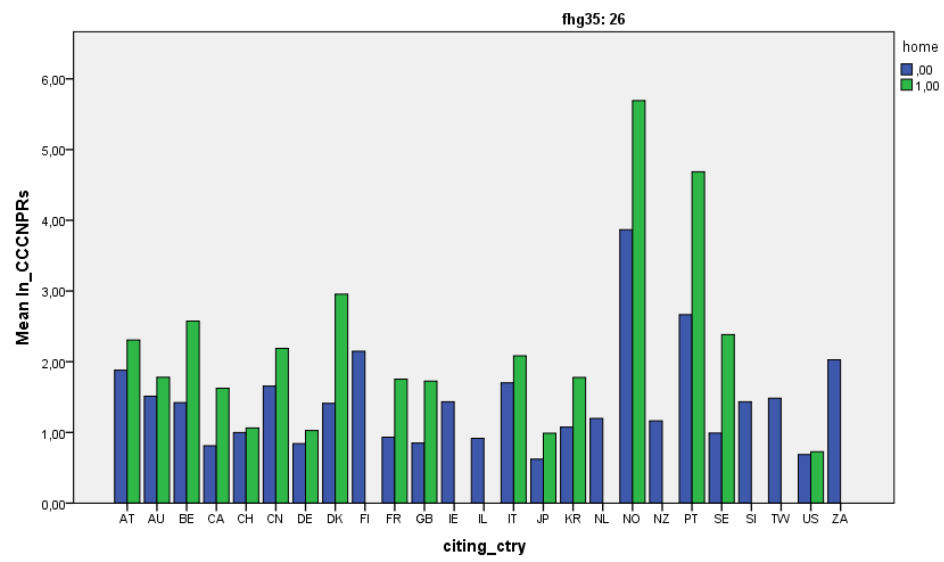
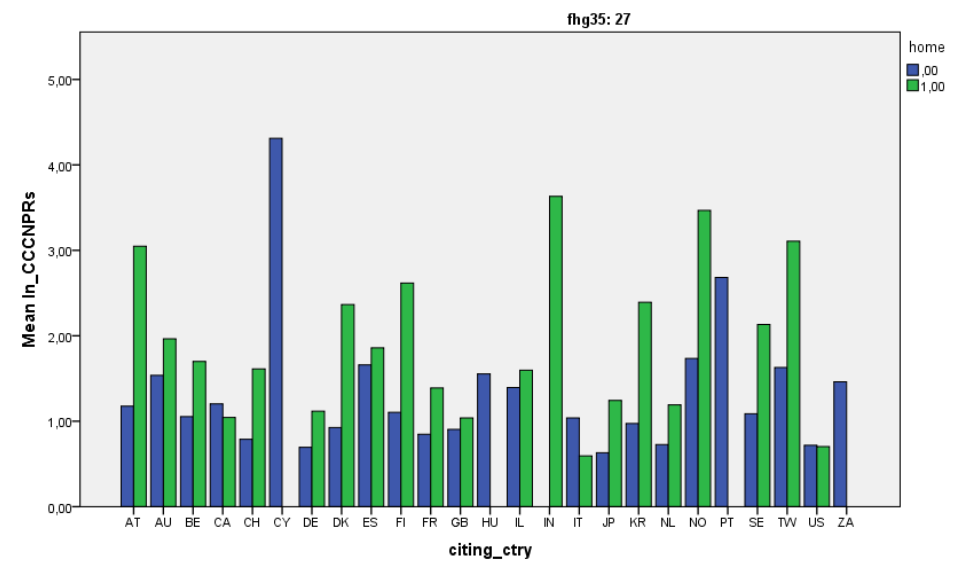
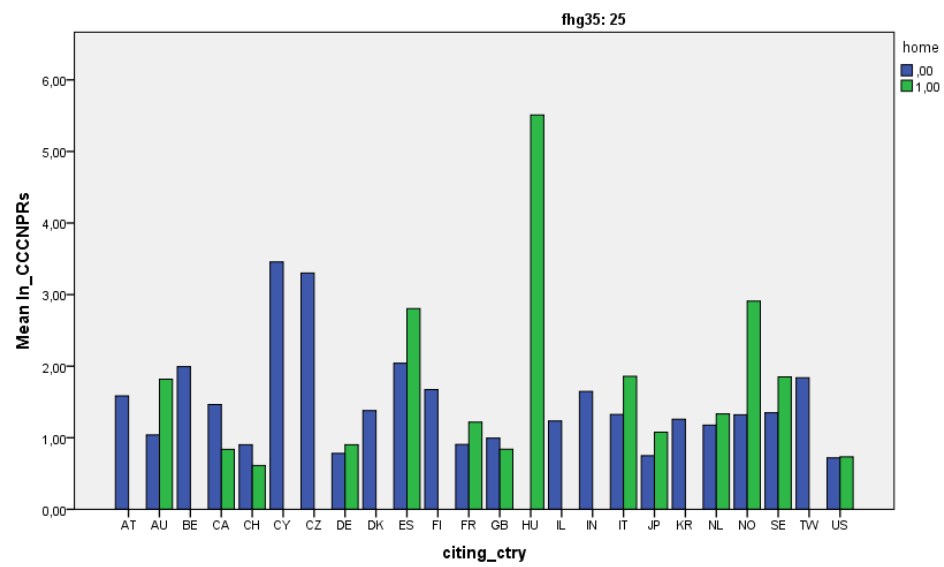




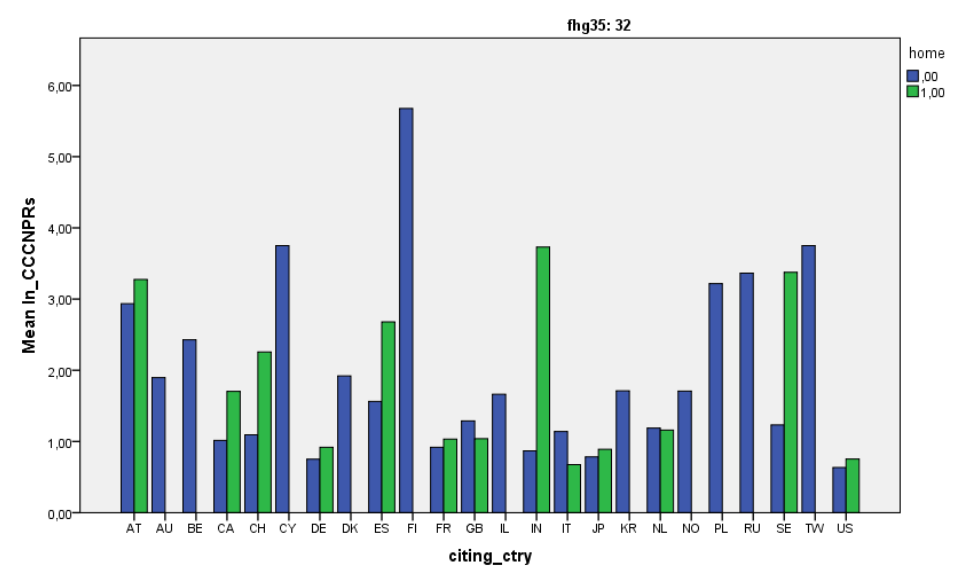
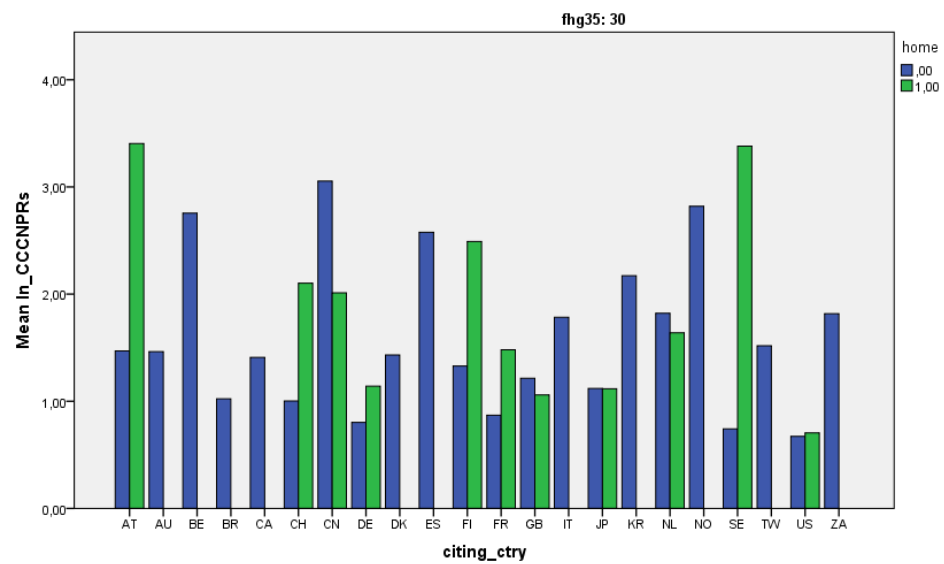
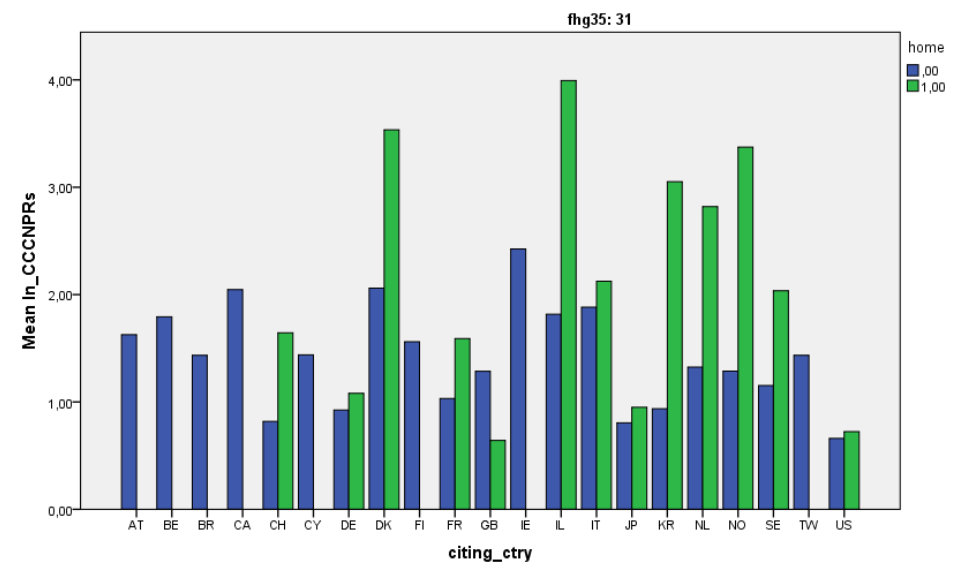
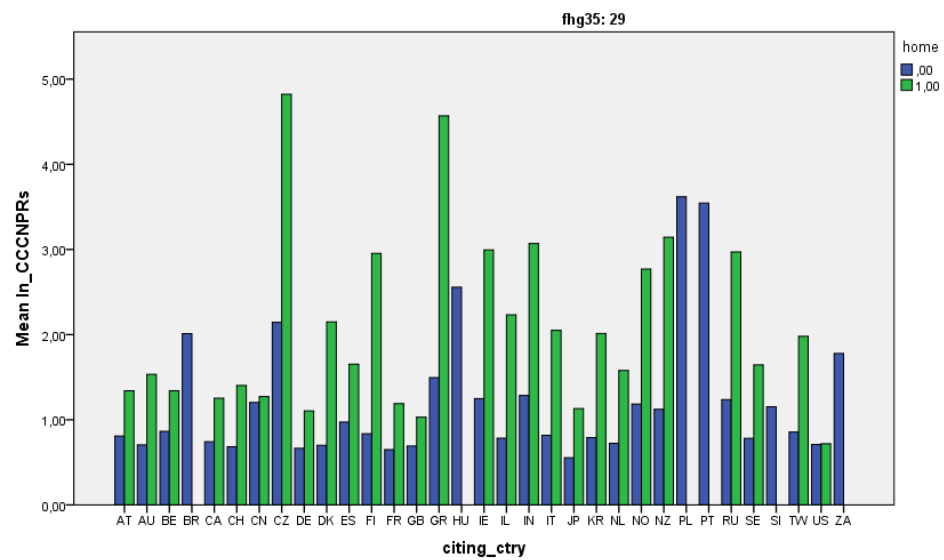


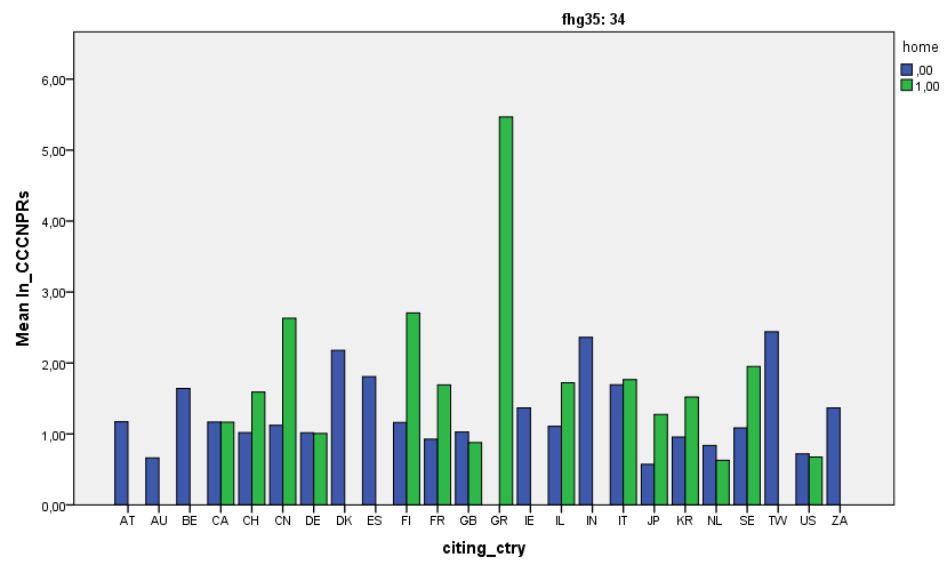
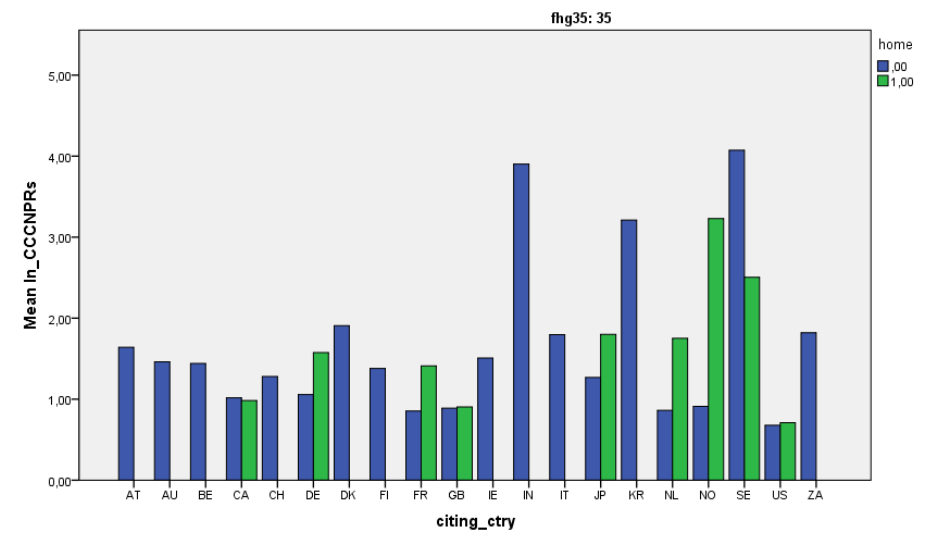
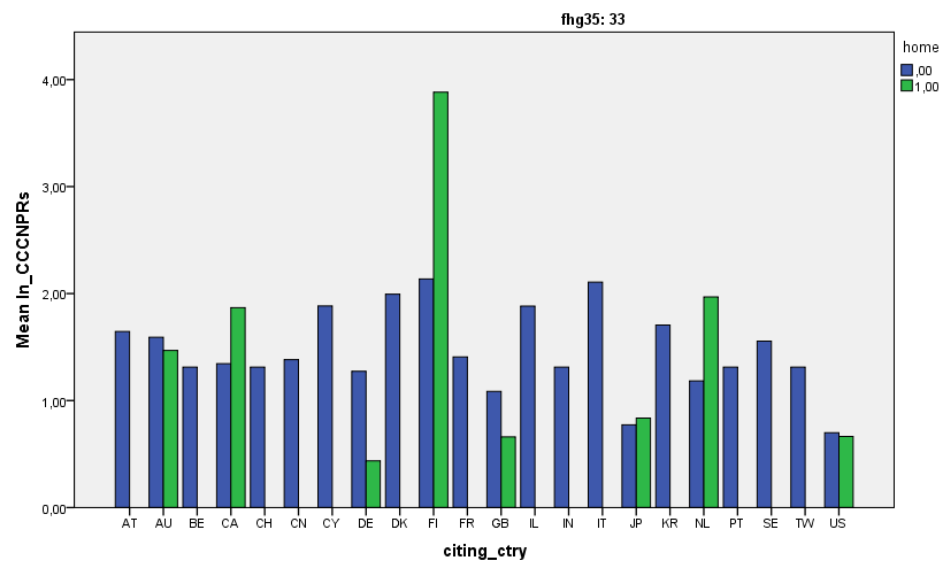












## ANNEX 2 – DETAILED ACCOUNT OF WoS-NPR MATCHING APPROACH (UNDELRYING FIGURES AND MATCHING RESULTS)

### **Step 1. Calculation of field-based match scores**

The method starts from match scores, calculated for a selection of seven fields that – in varying combinations – allow for the unique identification of individual scientific source documents:

PY (publication year),  
AUTHOR\_LN (last name of first author),  
SO (journal title),  
VL (volume),  
ISSUE (issue),  
BP (beginning page),  
TI (article title)

Each of these fields is available in the Web of Science, in parsed format.

Match scores between a WoS-field and an NPR string are calculated as: *the number of distinct terms in the WOS field that occur also in the NPR text-string, divided by the number of distinct terms in the WOS-field.*

Terms are identified as any string of characters enclosed by a non-numerical or non-alphabetical character. In practice, this means that a term is defined as a string of alphanumerical characters delimited by space or any punctuation character (e.g. comma, point, hyphen, semicolon). Hence strings containing both digits and non-digits will be regarded as one term, and strings with only digits will only be matched with other strings if the complete term is identical. E.g., a term containing “4” in an NPR-string (e.g. denoting beginning page) will not be matched with a term “14” in a WoS field, nor with a term “PH4” (e.g. part of a chemical formula).

Due to the sizes of both databases between which the matching is done, the required computing and data storage capacity becomes considerable (for more details regarding strategies to optimize computation time, see report deliverable 1.5 March 2012). In order to keep the required computing capacity within manageable limits, an a priori filter was set, based on publication years, narrowing down the number of WoS-documents that have to be considered for every NPR. More specifically, starting from all NPRs present in patents within a certain application year, a filter is set on those NPRs that contain at least one 4-digit number between 1987 and 2011. These 4-digit numbers are considered as potential publication years of the cited scientific reference. After the identification of these potential publication dates in the NPRs, match scores are calculated for each NPR, paired with all

WoS documents within a 5-year time window (publication year plus / minus 2 years). Take the following NPR as an example:

*DERRYBERRY R.T. ET AL.: 'On CDMA 2000 evolution - Reverse high-speed packet data physical layer enhancements in CDMA 2000 1xEV-DV' IEEE COMMUNICATIONS MAGAZINE vol. 43, no. 4, 01 April 2005, pages 41 - 47, XP001228792*

This NPR contains two 4-digit numbers between 1987 and 2011, namely "2000" and "2005". It will therefore be matched only to the subsets of WoS documents with a publication year between 1998-2002 and with a publication year between 2003-2007. This a priori publication year filter reduces the amount of WoS documents to be considered for every NPR by a factor 6.

## **Step 2. Download of NPR-WoS pairs with match scores above specified thresholds**

Before downloading the resulting matching scores of NPR-WoS combinations for validation and classification, two more filters were set to reduce the amount of data to be processed. Because of the large size of both the NPR dataset and WoS dataset, it is practically impossible to store all matching scores for all NPR-WoS combinations. Hence, a first filter was set to eliminate all combinations that are definitely not a match and to only store matching scores of potentially correct matches. The definition of this filter is based on the fact that it requires multiple field matches to results in an overall NPR-WoS match, and that the value of some fields is more important in the matching process compared to other fields. In a previous matching exercise (see Verbeek et al., 2002<sup>13</sup>), a large number of potential NPR-WoS combinations was validated and this validation made clear that no reliable matches were found for combinations that do not have a match on at least 4 fields (out of 5: publication year, author, volume, issue and beginning page), which can be reduced to a match on at least 3 fields under the condition that the author field does match (so: a match on author and at least two other fields), or if a substantial match on title exists (so: substantial match on title and at least three other fields). For our current matching, we lowered these thresholds and defined the following first raw filter for deciding which NPR-WoS combinations to store for further validation and classification:

### *1<sup>ST</sup> (RAW) LAYER FILTER*

\*\*\* (Sum of scores for the fields PY, AUTHOR\_LN, VOLUME, ISSUE, BP >= 3.5)

**OR**

(Sum of scores for the fields PY, AUTHOR\_LN, VOLUME, ISSUE, BP >= 3.0 **AND** at least 1 partial match on AUTHOR\_LN (i.e. AUTHOR\_LN > 0.5))

**OR**

(Sum of scores for the fields PY, AUTHOR\_LN, VOLUME, ISSUE, BP >= 2.5 **AND** at least 1 substantial match on TI (TI >= 0.7 and TI contains at least 4 distinct terms)))\*\*\*

---

<sup>13</sup> Verbeek, A., Debackere, K., Luwel, M., Andries, P., Zimmermann, E., & Deleus, F. (2002). Linking science to technology: Using bibliographic references in patents to build linkage schemes. *Scientometrics*, 54(3), 399–420.

This first raw filter still results in a huge amount of NPR-WoS combinations to be considered for classification and validation (e.g. for all 452.866 NPR's present in patents filed in 2007 and containing a potential publication year between 1987 and 2011, more than 67 million NPR-WoS combinations were retained after this first raw filter). There is no doubt that the vast majority of those combinations retained after the first raw filter are no matches. As we are, in this phase of the process, particularly interested in the potential matches, a second filter was set to further reduce the amount of data to be considered by identifying potential matches. The definition of this filter is again based on our previous validation, revealing that the vast majority of reliable NPR-WoS matches are combinations that do have a match on at least 4 fields (out of 5: publication year, author, volume, issue and beginning page), including the author field, or with a substantial match on title or journal. For our current matching, we lowered these thresholds and defined the following second filter to identify potential matches:

## *2<sup>ND</sup> LAYER FILTER*

\*\*\* (A full or partial match on the following 5 fields: PY, AUTHOR\_LN, VOLUME, ISSUE, BP)

**OR**

(A full or partial match on 4 out of the following 5 fields: PY, AUTHOR\_LN, VOLUME, ISSUE, BP

**AND**

(partial match on AUTHOR\_LN (AUTHOR\_LN  $\geq$  0.5))

**OR**

(partial match on TI (TI  $\geq$  0.4) **AND** TI contains at least 3 distinct terms)

**OR**

(partial match on SO (SO  $\geq$  0.6)))

**OR**

(A full or partial match on 3 out of the following 5 fields: PY, AUTHOR\_LN, VOLUME, ISSUE, BP

**AND** a partial match on AUTHOR\_LN (AUTHOR\_LN  $\geq$  0.6)

**AND** (a partial match on TI (TI  $\geq$  0.5) **AND** TI contains at least 3 distinct terms)

**OR**

(partial match on SO (SO  $\geq$  0.6)))\*\*\*

Table 2.1. shows the considerable reduction in the number of potentially matchable NPRs that is caused by these two filters.

Application year	total # NPRs	# NPRs after initial 2 filters	% reduction
1993	352.816	46.426	87%
1994	460.503	85.839	81%
1995	723.368	157.986	78%
1996	532.667	142.728	73%
1997	674.315	199.290	70%
1998	703.119	232.285	67%
1999	812.603	292.806	64%
2000	965.667	371.207	62%
2001	1.027.899	414.389	60%
2002	1.107.040	479.656	57%
2003	1.066.291	462.135	57%
2004	1.105.543	486.554	56%
2005	1.026.694	452.833	56%
2006	836.612	367.681	56%
2007	614.495	271.395	56%
2008	378.668	181.931	52%
2009	115.982	76.809	34%

**Table 2.1: Reduction in matchable NPRs in Patstat 2010-09 applications with application years 1993-2009 caused by 1<sup>st</sup> and 2<sup>nd</sup> layer filters**

### ***Step 3. Exploration and validation of additional filters on match scores to detect 'certain' matches***

The thresholds from step 2 were set sufficiently low to avoid a priori elimination of correctly matched pairs, i.e. to assure recall for those NPRs that are matchable. This implies that the resulting matches with the remaining NPRs will contain a considerable proportion of false positives. In a third step therefore, the match scores for the matched WoS-NPR pairs were evaluated and validated for being able to identify thresholds beyond which the resulting pairs are certain matches, with the aim of eliminating false positives and assuring precision. This validation was performed on the NPR sets from application years 2003 and 2007. Several filters were tested, using different combinations of match scores for the considered fields. For each filter, random sets of resulting pairs ( $N = 100$  à  $300$ ) were manually validated and – based on the results of these quality controls – a decision was made on whether or not to withhold the filter. The latter decision was based primarily on the precision of the matching results, while at the same time, the volume of retrieved matches needed to be sufficiently high for withholding the filter.

The filtering was performed in sequence, meaning that each filter was set on the set of matches that remained after the previous filters. After several iteration rounds, the following filters were withheld:

FILTER 1	MATCH SCORE TITLE $\geq 0,9$	
FILTER 2	(excluding matches from FILTER1) MATCH SCORE AUTHOR = 1 MATCH SCORE TITLE $\geq 0,7$	<b>AND</b>
FILTER 3	(excluding matches from FILTERS 1 and 2) MATCH SCORE SOURCE $\geq 0,9$ MATCH SCORE AUTHOR = 1 MATCH SCORE VOLUME = 1 MATCH SCORE BEGINNING PAGE = 1	<b>AND</b> <b>AND</b> <b>AND</b>
FILTER 4	(excluding matches from FILTERS 1, 2 and 3) MATCH SCORE AUTHOR = 1 MATCH SCORE VOLUME = 1 MATCH SCORE BEGINNING PAGE = 1 LENGTH OF NPR TEXT STRING $< 86$ CLASS NPR = 1	<b>AND</b> <b>AND</b> <b>AND</b> <b>AND</b>

FILTER 4 deserves some further explanation. It is primarily based on the observation that many NPRs do not contain the title of the cited document. At the same time, the unique identification of the WoS source document is feasible when based only on Author, Volume and Beginning page. However, NPRs containing an article title yielded a considerable volume of false positives with a filter including only these 3 fields. This is mostly due to the matching of digits occurring in the NPR title (whereby the matching wrongfully consider these digits to be the WoS volume or WoS beginning page that is sought for in the NPR). An example can serve to illustrate this point.

Consider the following NPR:

***Hansen*** et al., 529 km Unrepeated Transmission at **2.488** Gbit/s Using Dispersion Compensation, Forward Error Correction, and Remote Post- and Pre-amplifiers Pumped by Diode-Pumped Raman Lasers, *IEEE Electronics Letters*, Online No. 19951043, Jul. **7**, 1995. It is matched to the following WoS article (with MATCH SCORE AUTHOR = 1; MATCH SCORE VOLUME = 1 and MATCH SCORE BP = 1):

TI:	PARALLEL RENDERING
SO:	IEEE PARALLEL & DISTRIBUTED TECHNOLOGY
VL:	2
IS:	2
BP:	7
PY:	1994
AUTHOR:	HANSEN, C

The Author match is correct (although it may be a homonym). The WoS volume (2) matches to the underlined bold '2' digit that appears in the title part of the NPR (extracted from "2.488 GBit"). The beginning page matches to a part of the date mentioned in the WoS ("Jul.7). The resulting match is clearly a false hit.

If we cut off the length of the NPR string to a maximum of 85 characters, most remaining NPRs look like the following example:

*Cribbs et al. (1998). Circ Res 83:103-109.*

...which was correctly matched to the following WoS article:

TI: CLONING AND CHARACTERIZATION OF ALPHA 1H FROM HUMAN HEART, A  
MEMBER OF THE T-TYPE CA2+ CHANNEL GENE FAMILY  
SO: CIRCULATION RESEARCH  
VL: 83  
IS: 1  
BP: 103  
PY: 1998  
AUTHOR: CRIBBS, LL

Hence, for NPRs of these formats (which is the case for the lion share of short NPRs), confusion with digits other than volume or beginning page becomes much less likely, as has been confirmed by further validation efforts.

Therefore, the filter additionally required a criterion to delineate NPRs without article titles. An exploration and visual scan of NPR lengths revealed that strings below 86 characters were NPRs without article titles and of the format in the above example.

A final addition to FILTER 4 was required as visual inspection revealed that a considerable portion of these short NPR strings actually refer to patent-related documents, containing several digits that lead to false hits. An example of such NPRs is:

*U.S. Appl. No. 11/167,828, filed Jun. 27, 2005, Li.*

...which was wrongfully matched to the following WoS article:

TI: INFLUENCE OF POLARIZATION ON PROPERTIES OF 0-3 CEMENT-BASED PZT  
COMPOSITES  
SO: CEMENT & CONCRETE COMPOSITES  
VL: 27  
IS: 1  
BP: 27  
PY: 2005  
AUTHOR: LI, ZJ



For eliminating those NPRs that refer to patent documents and non-scientific documents, we relied on the NPR identification method developed by Callaert et al. (2012) and referred to above.

As stated earlier, filters were only withheld if the resulting pairs were correct matches (avoiding false positives) and of a sufficiently high volume. For checking the precision of the filters, validation efforts were performed on randomly chosen samples of the resulting NPR-WoS pairs for each filter for the 2003 and 2007 NPR subsets (sample results: see report March 2012 deliverable 1.5).

***Extraction of the 'certain' NPR-WoS pairs, according to the validated criteria specified in step 3.***

Table 2.2 gives an overview of the matching results (i.e. the number of retrieved pairs and the number of matched unique NPRs) for each of the 4 filters, broken down by the application year subsets.

Appl Year	# unique NPRs	# unique WOS	# resulting (non-filtered) pairs	# pairs resulting from FILTER 1	# unique NPRs	# pairs resulting from FILTER 2	# unique NPRs	# pairs resulting from FILTER 3	# unique NPRs	# pairs resulting from FILTER 4	# unique NPRs	TOTAL # FILTERED PAIRS	# unique NPRs
'93	46426	83299	173329	21569	21468	4294	4226	3473	3327	3138	3085	32474	31738
'94	85839	137770	310751	43322	43117	8273	8139	6497	6269	6802	6608	64894	63445
'95	157986	208692	579058	82981	82597	15299	15029	12533	12034	13617	13265	124430	121589
'96	142728	239171	601195	75784	75310	13284	13054	9048	8794	9707	9426	107823	105283
'97	199290	323266	896520	108921	108320	18065	17755	12208	11855	12456	12040	151650	148210
'98	232285	384828	1174676	129793	128883	20266	19862	13024	12672	13687	13299	176770	172602
'99	292806	484755	1692215	164419	163180	24425	23912	16304	15914	17244	16667	222392	216936
'00	371207	620291	2579996	206018	204330	28947	28348	17882	17380	20532	19825	273379	266506
'01	414389	722328	3294851	228255	226272	31528	30918	19453	18940	20322	19560	299558	291999
'02	479656	786063	4059479	273285	270811	34582	33830	20153	19496	21510	20570	349530	340019
'03	462135	865329	4191940	251236	249217	30932	30258	20010	19515	20612	19750	322790	314649
'04	486554	929547	4710235	258072	255936	31192	30532	19006	18629	18903	18105	327173	319131
'05	452833	961815	4671606	243474	241495	28691	28062	16424	16117	17678	16971	306267	298568
'06	367681	899195	4187639	201374	199597	22112	21589	11731	11519	12252	11793	247469	241328
'07	271395	832586	3612443	148019	146673	15807	15418	10085	9879	11918	11458	185829	181002
'08	181931	749020	3276270	100443	99486	9287	9124	7768	7644	11046	10577	128544	125196
'09	76.809	534127	1993799	47471	4.971	3890	3817	1934	1897	1749	1643	55044	53591

**Table 2.2: Matching results for withheld filters**

The observation that the number of resulting pairs is higher than the number of unique NPRs is due to the fact that some NPRs are matched to more than one WoS record. This is the case for about 2% of the matched NPRs (see annex 2). A further analyses of these cases show that in many cases, the multiple matches are in reality accurate matches; with some WoS documents appearing as different document types in the Web of Science, but basically referring to the same scientific document. For an illustration with an example, we refer to the methodological project report dating from March 2012.

After criteria setting, validation and quality checks of the extracted NPR-WoS pairs, Table 2.3 shows the proportions of unique NPRs for which the matching was successful, i.e. the proportions of NPRs for which a unique WoS source document was identified, by application year subset. These pairs will serve as the basis for the indicator development in Deliv 1.5. Proportions are calculated on the volumes of unique NPRs, the denominator being the starting set of NPRs after the filters / thresholds set in step 2 (cf. supra). It can be seen that between 66% and 70% of matchable NPRs are covered with the defined filters.

<b>application year</b>	<b># unique NPRs</b>	<b>% unique NPRs matched in FILTER 1</b>	<b>% unique NPRs matched in FILTER 2</b>	<b>% unique NPRs matched in FILTER 3</b>	<b>% unique NPRs matched in FILTER 4</b>	<b>TOTAL % MATCHED UNIQUE NPRs</b>
<b>1993</b>	46.426	46%	9%	7%	7%	68%
<b>1994</b>	85.839	50%	10%	7%	8%	74%
<b>1995</b>	157.986	52%	10%	8%	9%	77%
<b>1996</b>	142.728	53%	9%	6%	7%	74%
<b>1997</b>	199.290	54%	9%	6%	6%	74%
<b>1998</b>	232.285	55%	9%	5%	6%	74%
<b>1999</b>	292.806	56%	8%	5%	6%	74%
<b>2000</b>	371.207	55%	8%	5%	6%	72%
<b>2001</b>	414.389	55%	8%	5%	5%	70%
<b>2002</b>	479.656	56%	7%	4%	4%	71%
<b>2003</b>	462.135	54%	7%	4%	4%	68%
<b>2004</b>	486.554	53%	6%	4%	4%	66%
<b>2005</b>	452.833	53%	6%	4%	4%	66%
<b>2006</b>	367.681	54%	6%	3%	3%	66%
<b>2007</b>	271.395	54%	6%	4%	4%	67%
<b>2008</b>	181.931	55%	5%	4%	6%	69%
<b>2009</b>	76.809	61%	5%	2%	2%	70%

**Table 2.3: Matching results: proportions of successfully matched unique NPRs**

# **ANNEX 3 – GRAPHS SUPPLY VERSUS ABSORPTION OF SCIENCE PER CITING COUNTRY, BY PATENT OFFICE**

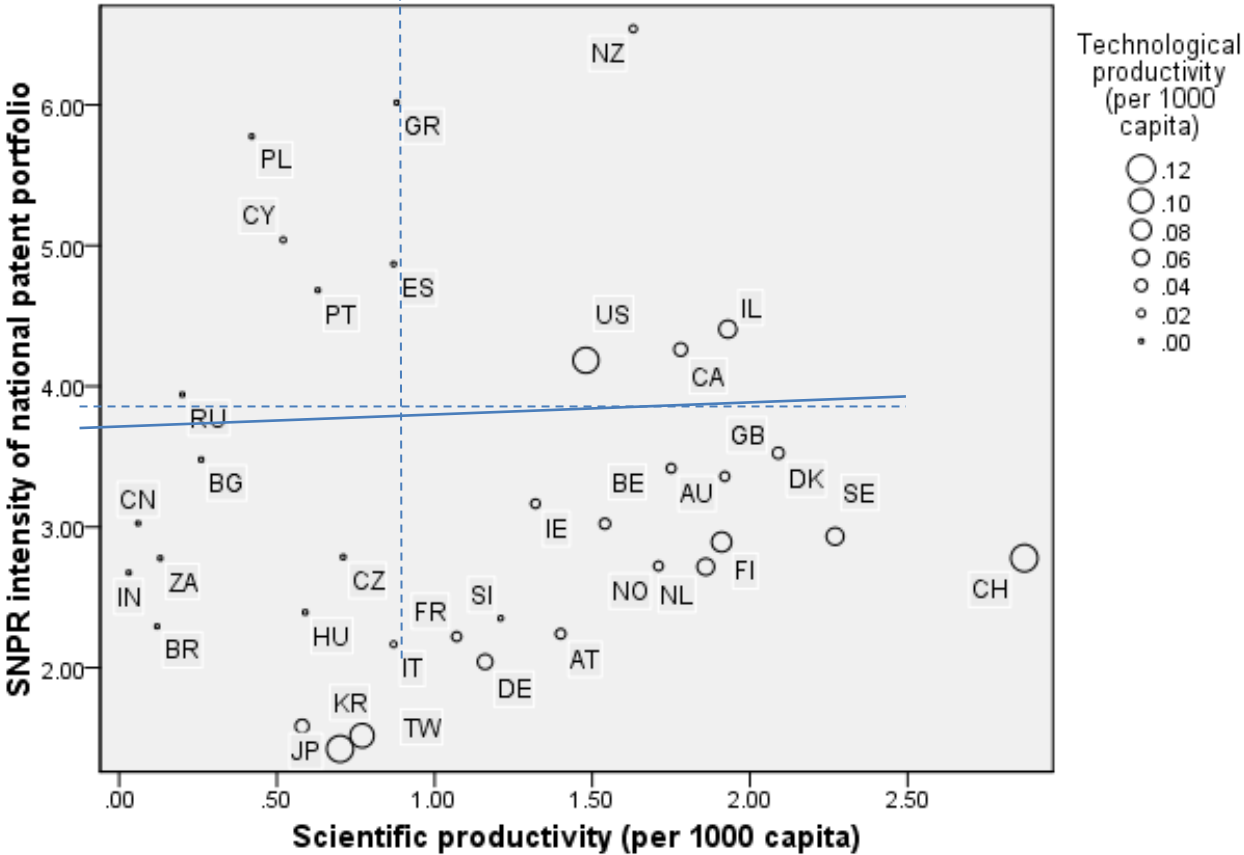


Figure 2a: Supply (1991-2011) versus absorption of science per citing country (2000-2009) – USPTO patents.

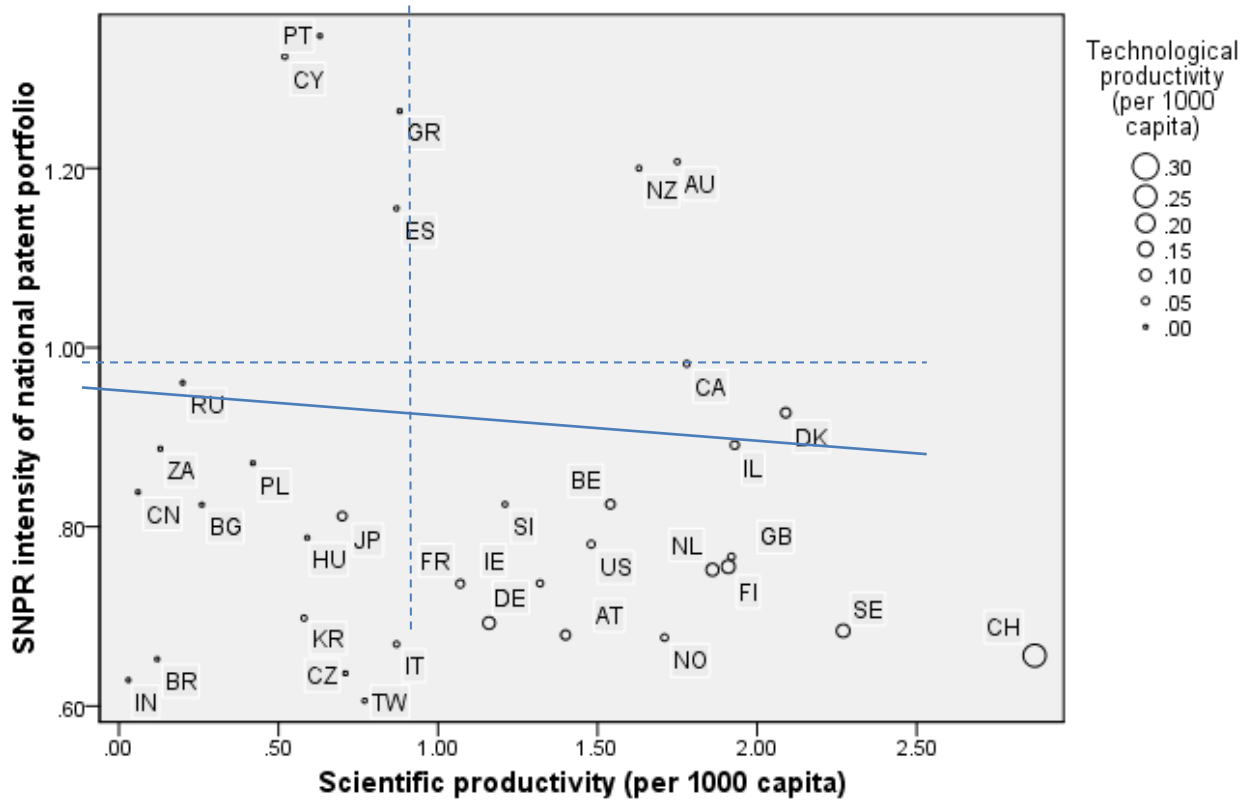
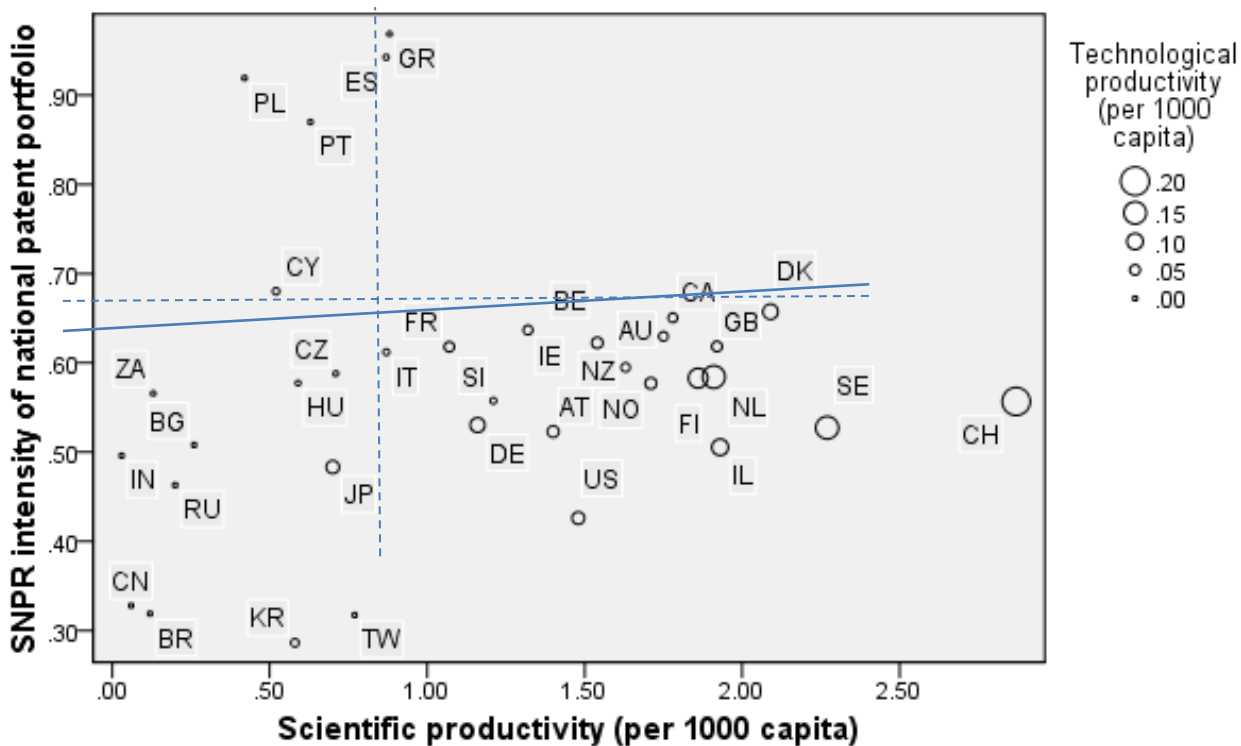


Figure 2b: Supply (1991-2011) versus absorption of science per citing country (2000-2009) – EPO patents.



**Figure 2c: Supply (1991-2011) versus absorption of science per citing country (2000-2009) – PCT patents.**

European Commission

**Patterns of Science-Technology Linkages**

Luxembourg: Publications Office of the European Union

2014 — 100 pp — 21 x 29,7 cm

ISBN 978-92-79-34793-1

doi:10.2777/55249





## How to obtain EU publications

### Free publications:

- one copy:  
via EU Bookshop (<http://bookshop.europa.eu>);
- more than one copy or posters/maps:  
from the European Union's representations ([http://ec.europa.eu/represent\\_en.htm](http://ec.europa.eu/represent_en.htm));  
from the delegations in non-EU countries ([http://eeas.europa.eu/delegations/index\\_en.htm](http://eeas.europa.eu/delegations/index_en.htm));  
by contacting the Europe Direct service ([http://europa.eu/eurodirect/index\\_en.htm](http://europa.eu/eurodirect/index_en.htm)) or  
calling 00 800 6 7 8 9 10 11 (freephone number from anywhere in the EU) (\*).  
(\* ) The information given is free, as are most calls (though some operators, phone boxes or hotels may charge you).

### Priced publications:

- via EU Bookshop (<http://bookshop.europa.eu>).

### Priced subscriptions:

- via one of the sales agents of the Publications Office of the European Union  
([http://publications.europa.eu/others/agents/index\\_en.htm](http://publications.europa.eu/others/agents/index_en.htm)).

In this report, the relationships between science and technology are mapped and analyzed by means of citations from patent documents (EPO, USPTO and PCT) towards scientific publications (Web of Science).

Focusing on national innovation systems (countries), the findings reveal the presence of a home 'bias': for the majority of the countries under study, science situated within the boundaries of the country is more often cited than foreign science. In terms of performance, it becomes apparent that a broader geographical scope in terms of scientific prior art is associated with higher levels of patent activity.

At the same time, our findings also highlight a strong positive relationship between scientific capabilities (of countries) and their technological performance.

#### *Studies and reports*

